Statistical Methods in Financial Risk Management

Rodrigo dos Santos Targino

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University College London.

Department of Statistical Science
University College London

April 19, 2017

Declaration of authorship

I, Rodrigo dos Santos Targino, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This thesis studies two different problems regarding financial companies' capital, which is a buffer used to cover unexpected losses. The first problem is related to *understanding* the capital, in a process called *allocation* and the second is connected to its *management*.

For the capital allocation problem we follow the Euler principle and develop simulation-based algorithms to efficiently compute the contribution of individual names to the portfolio's overall capital. Although the algorithms proposed in this thesis are general enough to be applied to any portfolio, we focus on the allocation of operational risk and insurance capital. In both cases the algorithms proposed in this thesis are based on Sequential Monte Carlo (SMC) methods. In the context of operational risk we assume annual losses in the business lines are jointly modelled through a copula, and no parameter uncertainty is involved. In this case we are able to use the copula dependence structure to devise efficient algorithms. For the allocation of the one-year reserve risk and the one-year premium risk of insurance companies we develop a novel and fully Bayesian claims reserving model, and discuss how to perform allocations under parameter uncertainty.

Further to understanding the company's capital we develop a class of financial instruments to facilitate the transference of operational risks, which would naturally lead to capital reductions. As the annual amount due to operational losses can be extremely large "full insurance coverage" is very expensive, preventing some companies from accessing the insurance market. To circumvent this problem we propose a class of insurance products that last for T years but the policyholder is only allowed to make claims for insurance coverage in $k \leq T$ years. For some combinations of annual coverage and loss distribution we are able to derive the optimal usage strategy for these products in closed form and for general cases we present an approximation scheme based on density series expansion.

Acknowledgements

I would like to thank my supervisor, Gareth W. Peters, for all the support throughout the past four years. Since the very first day Gareth has done everything to ensure I enjoyed my research as much as he enjoys his. I am also grateful to my co-supervisors, Pavel V. Shevchenko and Mario V. Wüthrich for their hospitality at, respectively, CSIRO and ETH, and for everything I have learned from them.

It was a pleasure to study alongside great researchers and friends at the Department of Statistical Science at UCL. My weekdays, Fridays in particular!, were extremely enjoyable due to the company of my fellow PhD students. Without them I would not have coped with a window-less room for four years!

From the other side of the Atlantic, in Brazil, I received the support of my friends and also from the generous Ciência sem Fronteiras scholarship, provided by the Conselho Nacional de Pesquisa e Desenvolvimento (CNPq).

Last but, obviously, not least, I thank my wife Bibiana for her love, care and patience during my PhD.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to financial risk management

The importance of financial risk management has been progressively increasing for all corporations in the last few decades, notably in financial institutions. The first question one should pose is, then, *what exactly is a financial risk?* Although dictionaries would define risk as "the possibility of something bad happening" or "a situation involving exposure to danger", in the context of financial risks the definition provided in [McNeil et al., 2010] is perhaps the most adequate one:

> Financial risk is *the quantifiable likelihood of loss or less-than-expected returns.*

In a quantitative framework, the notion of risk is invariably related to *uncertainty* and, therefore, to *randomness*. The last concept has been formally defined in the first decades of the twentieth century with the axiomatization of the probability theory by A. N. Kolmogorov. As this thesis focuses on a quantitative approach to risk management we rely on the theories of probability and statistics.

Within the context of finance and insurance, three main risk classes can be characterized: *market risk*, *credit risk* and *operational risk*. The first risk, undoubtedly the one for which most of the attention has been drawn on the twentieth century, is related to "losses in positions arising from movements in market prices", [BCBS, 2003b]. *Credit risk*, in turn, is the risk of not receiving agreed payments due to a default of the borrower or, more formally "the risk that a counterparty will not settle an obligation at its agreed full value, either when due or at any time thereafter", [BCBS, 2003b]. Although *market* and *credit* losses are not desired, due to the nature of banking/insurance business, they can be seen as *expected*. On the other hand *operational losses*, are comprised of a combination of both expected and unexpected losses and and comprise all the losses resulting from "inadequate or failed internal processes, people and systems, or from external events", [BCBS, 2006]. From a quantitative point of view, another important notion that is universal in any quantification of risk management is that now known as *model risk*, which is associated with using an inappropriate model to measure risk. For example, one would be underestimating the risk of large losses if modelling the loss distribution

with lighter tails than the true (unobservable) distribution.

The definition of risk itself may already indicate why companies invest in measuring and understanding risk and loss processes but it is important to note that "earnings stability and the survival of the company are important managerial objectives", [Hull, 2012]. Apart from the managerial objectives, risk management in financial institutions is also a regulatory requirement, as discussed in the next section.

1.1 Regulatory issues: Basel and Solvency accords

Unlike many other sectors of the economy, the financial sector, throughout the world, is heavily (and increasingly) regulated. Governments and their regulatory agencies want the sector to be as stable as possible, both locally (in their own countries) and globally, in an attempt to avoid repeated cycles of financial crisis and recessions/depressions of economies, which usually end up with expensive governmental bailouts (see [Reinhart and Rogoff, 2009] for a historical account on financial crisis). To ensure companies are going to be solvent in a finite horizon (of usually one year), regulators require them to set aside some *capital*, which should be seen as a buffer to cover for unexpected losses.

As per [McNeil et al., 2015], "the main aim of modern prudential regulation has been to ensure that financial institutions have enough capital to withstand financial shocks and remain solvent. Robert Jenkins, a member of the Financial Policy Committee of the Bank of England, was quoted in the *Independent* on 27 April 2012 as saying: *Capital is there to absorb losses from risks we understand and risks we may not understand. Evidence suggests that neither risk-takers nor their regulators fully understand the risks that banks sometimes take. That's why banks need an appropriate level of loss absorbing equity.*"

In this section we provide an overview of how regulation in financial/insurance market has evolved since the 1970's. The reader is referred to [BCBS, 2003a] for the history of the Basel Committee and to [Tarullo, 2008] for further details.

1.1.1 Banking regulation and the Basel Accords

The first modern attempt to design international regulatory standards in the banking industry dates back to 1974, when, in a cross-jurisdictional event, West Germany's Federal Banking Supervisory Office forced the liquidation of the Bank Herstatt on the 26th of June. This event was followed by the bankruptcy of the Franklin National Bank of New York, in October and both are related to the breakdown (in the early 1970's) of the Bretton Woods system of managed exchange rates.

As a response to these incidents, the Group of Ten (see Table 1.1) established, late in 1974, the *Committee on Banking Regulations and Supervisory Practices* – to be later renamed to its current name, the *Basel Committee on Banking Supervision* (BCBS). Since 2009 the Committee is composed of 27 international bodies (see [BCBS, 2003a, Appendix A]).

As stated in [BCBS, 2003a], the Committee's *aim was and is to enhance financial stability*

Belgium	Netherlands
Canada	Sweden
France	Switzerland*
Germany	United Kingdom
Italy	United States
Japan	

Table 1.1: Composition of the Group of Ten (G-10) in 1974: original members (since inception, in 1962) and Switzerland (joined in 1964).

by improving supervisory knowhow and the quality of banking supervision worldwide. Even though the Committee does not possess any formal supranational authority, its supervisory standards are expected to be followed by the national supervisory authorities, with any necessary adjustments to the local jurisdiction. From a legal point of view, local authorities, e.g., Central Banks, are responsible for legally enforcing compliance with the guidelines set by the Committee.

1.1.1.1 Basel I: the Basel Capital Accord

Differently from the 1974 events whose origins trace back to European *market risk* events, in the 1980's the world would see a *credit risk* crisis in Latin America. The "lost decade" of 1980 in Latin America was preceded by boom years in the 1960's and 1970's, when local military dictatorships increased sovereign debts.

The Latin American debt crisis led the BCBS to develop the Basel Accord of 1988 (Basel I), whose main emphasis was on credit risk. This accord took an important step towards an international minimum capital standard, where the *capital requirement* is to be understood as the amount of capital a financial institution is required to hold by it local regulator.

The 1988 Accord introduced a notion of simple *risk weights*, where different classes of assets have different risk weights, ranging from 0% (for low risk assets) to 100% (for high risk assets). It is important to stress that the weights are defined by the local regulator but government bonds as well as cash, for example, usually have 0% weight while mortgages have 50% weight. Other types of loans to customers, in general, are assumed to have 100% weight. The sum of all assets weighted by its risk weight leads to the so-called *Risk Weighted Assets* (RWA). An example of the risk-weighting process (from [Cruz et al., 2015]) can be seen in Table 1.2

Another concept introduced in the first Basel accord was the classification of the capital into Tier 1 and Tier 2, as described in [BCBS, 1988, Annex 1]:

- Tier 1 (core capital)
  (a) Paid-up share capital/common stock;
  (b) Disclosed reserves.
- Tier 2 (supplementary capital):

Risk weight (%)	Asset	Amount ($)	RWA ($)
0	Cash	10	0
	Treasury bills	50	0
	Long-term treasury securities	100	0
20	Municipal bonds	20	4
	Items in collection	20	4
50	Residential mortgages	300	150
100	AA+ rated loan	20	20
	Commercial loans, AAA- rated	55	55
	Commercial loans, BB- rated	200	200
	Sovereign loans B- rated	200	200
	Fixed assets	50	50
Not rated	Reserve for loan losses	(10)	(10)
Total		1015	673

Table 1.2: Example of risk weighted assets calculation under Basel I (from [Cruz et al., 2015]).

(a) Undisclosed reserves;

(b) Asset revaluation reserves;

(c) General provisions/general loan-loss reserves;

(d) Hybrid (debt/equity) capital instruments;

(e) Subordinated debt.

The main requirement of the Basel I accord (implemented by the end of 1992) was that the *Capital Adequacy Ratio* (CAR) should be at least 8%, where the CAR (also known as Cooke ratio) is defined as the percentage of the institution's eligible capital (sum of Tier 1 and Tier 2 capital) to its RWA. In other words,

$$\text{CAR} = \frac{\text{Eligible capital}}{\text{RWA}} \geq 8\%. \tag{1.1}$$

Remark 1.1.1. *As mentioned in Paul Embrecht's presentation [Embrechts, 2008] the calculation of the capital ratio in (1.1) can be divided between "us" and "them". The denominator, which involves the risk weighted positions, is calculated by "us", the 'quants', while "they", the accountants, the management and the board are responsible by the numerator.*

1.1.1.2 Basel II: the New Capital Framework

One of the significant weaknesses of the 1988 Accord was that the *credit rating* of the borrower was not taken into consideration when calculating the risk weight of its debt.

To correct for this discrepancy and to include other features, in 1999, a decade after the release of the first Basel Accord the BCBS started the process to replace the 1988 Accord. The

result was the *Revised Capital Framework* or the *Basel II*, released in June 2004. This document comprised three *pillars*, namely,

1. Pillar 1 - Minimum capital requirements;
2. Pillar 2 - Supervisory review;
3. Pillar 3 - Market discipline.

In the first Pillar, the total capital did not change and banks were still required to hold at least 8% of the RWA, but the credit worthiness of the counterparts started to be reflected. While the original Basel I Accord only dealt with *credit risk* (with the 1996 Amendment including *market risk*) the Basel II Accord created a capital charge also for *operational risk*.

As noticed by [Hull, 2012], due to Pillar 2 "supervisors were required to do far more than just ensuring that the minimum capital required under Basel II is held". Some of the new attribution of the local regulators were, for example, to encourage banks to develop new risk management techniques and also to help financial institutions to evaluate risks not covered in Pillar 1.

The third Pillar required banks to publicly disclose the risk measures and any other information relevant to risk management, including how they allocate their capital. The allocation of capital is one of the main themes of this thesis, being discussed in Chapters 4 and 6. It has also been decided that the capital should be calculate as the Value at Risk (VaR) with a one-year time horizon and a 99.9% confidence level for operational risk.

To overcome banks' criticism about the coarseness of the risk weights from Basel I, in the new Accord banks were allowed to choose from three different approaches for handling credit risk:

1. The Standardized Approach;
2. The Foundation Internal Ratings Based (IRB) Approach;
3. The Advanced IRB Approach.

Even the most basic alternative (the Standardized Approach) already included some measures for better differentiation of risks through *credit ratings*, although the ratings themselves are not calculated internally when the bank chooses this approach. Larger and more complex banks were allowed to use Internal Ratings Based (IRB) approaches. In these cases the assessment of the riskiness of the credit portfolio could be done by the bank itself.

As mentioned in [McNeil et al., 2015], "a basic premise of Basel II was that the overall size of regulatory capital through the industry should stay unchanged under the new rules. Since the new rules for credit risk were likely to reduce the credit risk charge, this opened the door for operational risk". For the operational risk capital three approaches were introduced:

1. The Basic Indicator Approach (BIA);

Accord	Year	Key points
Basel I	1988	Introduces minimal capital requirements for the banking book. Introduces tier concept for capital requirement. Incorporates trading book into the framework later on through the Market Risk Amendment (MRA).
Basel II	2004	Allows usage of internal models and inputs in risk management. Introduces operational risk.
Basel II/III	2010	Increases capital requirement for trading book, with significant increase for correlation trading and securitization.
Basel III	2010	Motivated by the financial crisis of 2008, increases capital requirements, introduces leverage constraints and minimum liquidity and funding requirements.

Table 1.3: General summary of the Basel Accords (from [Cruz et al., 2015]).

2. The Standard Approach (SA);

3. The Advanced Measurement Approach (AMA).

As with the credit risk, the use of these approaches depend on the level of sophistication of the bank. Under the BIA the operational risk capital is set as the bank's average annual gross income over the last three years multiplied by 0.15. Under the SA approach there are different factors to be applied to the gross income from different business lines, this been the only difference from the BIA. In the AMA the bank is allowed to use its own internal models to calculate the operational risk capital. Another advantage of this approach is that the regulator can recognize the risk mitigation impact of insurance contracts (see Chapter 7).

1.1.1.3 Basel III

After an European trigger (1974) and a Latin American one (1980's) the 2000's witnessed an American born crisis (named by some as "The Crisis" [Das et al., 2013]), which, as usual, led to more regulation: this time the Basel III Accord. A summary of the key tekaways (compiled in [Cruz et al., 2015]) of the Basel Accords is found in Table 1.3.

The first proposals of the Basel III document were published in December 2009 but the final version was only available a year later, see [BCBS, 2010b] and [BCBS, 2010a] and its implementation will occur gradually between 2013 and 2019. As discussed in [Hull, 2012], there are six parts to this regulatory document:

1. Capital Definition and Requirements;

2. Capital Conservation Buffer;

3. Contercyclical Buffer;

4. Leverage Ratio;

5. Liquidity Risk;

6. Counterparty Credit Risk.

At the moment these lines are being written the supervisors at the BCBS are consulting the operational risk community on the possibility of scrapping the Advanced Measurement Approach (AMA), in a movement that started in October 2014. By that time the BCBS released the consultation document [BCBS, 2014] proposing a Revised Standardized Approach (RSA) for operational risk and in March 2016 the BCBS published the consultative document [BCBS, 2016b] which suggests the replacement of the AMA by a new non-model-based method, named Standardized Measurement Approach (SMA). The SMA is based on the combination of a simple standardized measure of operational risk, based on a fixed percentage of operating revenues, and bank-specific operational loss data (based on the arithmetic average of losses over the past 10 years). As this proposal discards the knowledge on operation risk modelling accumulated both by practitioners and academics it has been followed by a heated debate (see, e.g., [Peters et al., 2016a] and [Wills, 2016]).

1.1.2 Insurance regulation: Solvency I/II and the Swiss Solvency Test

Notwithstanding that banks have supranational regulatory standards, insurance companies, up to date, do not have any formal international regulation. In the United States insurance companies are regulated at the state level, with the national support of the National Association of Insurance Commissioners (NAIC). In Europe the European Union is in charge of the regulatory role, through the European Insurance and Occupational Pensions Authority (EIOPA, formerly known as CEIOPS: the Committee of European Insurance and Occupational Pensions Supervisors). On the other hand, since the 1st of January 2011, Switzerland has been using the *Swiss Solvency Test* (SST) for capital calculation, over-sighted by the Swiss Financial Markets Supervisory Authority (FINMA), a government body created in 2007 as a merge of the Federal Office of Private Insurance (FOPI), the Swiss Federal Banking Commission (EBK) and the Anti-Money Laundering Control Authority.

The recently replaced European regulatory framework, known as Solvency I, came to force in 2004 and was replaced by Solvency II (often called “Basel for insurers”) on the 1st of January 2016. While Solvency I calculated capital only for *underwriting risks*, the new directive has a much wider scope, considering, for example, operational risk capital. The informal name of the Solvency II directive is mainly due to the similarities it holds with Basel II. For example, under Solvency II there are also be three pillars, exactly as in Basel II.

Pillar 1 of Solvency II introduces two capital requirements, the Solvency Capital Requirement (SCR, discussed in Section 5.2.3) and the Minimum Capital Requirement (MCR). The SCR can be calculated using either a standard formula given by the regulators or an internal model developed by the insurance company. This is the capital required to ensure the company

	Business line
1	Corporate finance
2	Trading and sales
3	Retail banking
4	Commercial banking
5	Payment and settlement
6	Agency services
7	Asset management
8	Retail brokerage

	Event type
1	Internal fraud
2	External fraud
3	Employment practices and workplace safety
4	Clients, products and business practices
5	Damage to physical assets
6	Business disruption and system failures
7	Execution, delivery and process management

Table 1.4: Basel II business lines (left) and event types (right) – see [BCBS, 2006], Annexes 8 and 9.

will be able to meet its obligations over the next 12 months and if the capital falls bellow the SCR level the company should, at least, deliver a plan to the supervisor to restore its capital. Differently from Basel II, in the Solvency II regulation the capital involves the calculation of a Value at Risk (VaR) with 99.5% confidence (less than Basel II's 99.9% for OpRisk, for example), while the SST prescribes the calculation of the 99% Expected Shortfall (ES).

The MCR, which is intended to correspond to the $\text{VaR}_{85\%}$ (and is bounded between 25% and 45% of the SCR), can be regarded as a "hard" capital floor (while the SCR is a "soft" floor), a control level that, if breached, would trigger "ultimate supervisory action". In this case the "the insurer's liabilities will be transferred to another insurer and the license of the insurer will be withdrawn or the insurer will be closed to new business and its in-force business will be liquidated" (as stated in the European Commission MEMO/07/286).

1.2 Research questions and outline of the thesis

This thesis studies two different problems related to financial companies' capital. The first one is related to *understanding* the capital, in a process called *allocation*. In this regard we develop Sequential Monte Carlo (SMC) algorithms to compute the capital and also to break it down into the company's different constituents. The second problem is related to the capital *management*, where we develop insurance products to facilitate the transference of specific risks.

These problems are studied in two different contexts. The *capital allocation* problem is first presented as a way to understand the drivers of the capital related to operational risk, for example, distributing the capital amongst the combination of business lines and event types, as in Table 1.4. An idealized version of this problem is discussed in Chapter 4, when we first introduce a Sequential Monte Carlo algorithm to calculate, via simulation, the capital contributions. In this first instance we assume all the model parameters are perfectly known and focus solely on the allocation problem, which can be rewritten as an expectation conditional to a rare event. These results were published in [Targino et al., 2015].

Chapters 5 and 6 deal with the same allocation problem, in an actuarial context. Motivated by the short term view of the recent solvency regulations (such as the Swiss Solvency

Test and Solvency II), Chapter 5 is devoted to developing statistical models for the *one-year reserve risk* and the *one-year premium risk*. The later is constructed based on the Swiss Solvency Test directives and no parameter uncertainty is involved. For the former we extend the Bayesian gamma-gamma chain ladder model of [Gisler, 2006] and [Gisler and Wüthrich, 2008] and provide two distinct approximations to it, resulting in what we call the *marginalized* and *conditional* models. Both strategies approximate the Bayesian gamma-gamma model through log-normal distributions and matching of the first two moments, the difference being at which stage the approximation is performed. For the marginalized version we match the moments of a distribution where the unknown parameters have been integrated out, while in the conditional model we approximate the conditional distribution of the Bayesian gamma-gamma model by a log-normal.

In Chapter 6 we make use of the framework developed in Chapter 5 and present algorithms to solve the allocation problem under the marginalized and conditional approximations. As in the marginalized approach no parameter uncertainty is present (it is integrated out before the allocation process) the algorithm is mostly based on the one provided in Chapter 4. Allocations for the conditional model present an additional layer of complexity, as one needs to calculate conditional expectations with respect to a model whose density is not known in closed form. To overcome this difficulty we develop a *pseudo-marginal* SMC sampler.

Technical findings related to the extension of the Bayesian gamma-gamma chain ladder model are presented in [Peters et al., 2017] and its use in the capital allocation problem is detailed in [Peters et al., 2016b].

Chapter 7 returns the focus to operational risk modelling, and the aim is not to understand the capital anymore, but to construct instruments for the transference of risk, which would lead to capital reductions. In particular, we study a class of insurance products where the policy holder (say, a bank) has the option to insure k of its annual operational risk losses in a horizon of T years. This involves a choice of k out of T years in which to apply the insurance policy coverage by making claims against losses in the given year. Although this class of products can be used for mitigation of any risk, is particularly relevant for operational risk, due to the sheer scale of operational losses – which leads to expensive insurance products. As the buyer is only covered for k years (and not T) this type of product can substantially reduce insurance premiums, making it affordable to a larger proportion of companies.

The insurance product structure presented in Chapter 7 can accommodate any kind of annual mitigation, but we present two basic generic insurance policy structures that can be combined to create more complex types of coverage. Following the Loss Distributional Approach (LDA) with Poisson distributed annual loss frequencies and Inverse-Gaussian loss severities we are able to derive analytical expressions for the multiple optimal decision strategy that minimizes the expected operational risk loss over the next T years. For the cases where the combination of insurance policies and LDA model does not lead to closed form expressions for the multiple

optimal decision rules, we also develop a principled class of closed form approximations to the optimal decision rule. These approximations are developed based on a class of orthogonal Askey polynomial series basis expansion representations of the annual loss compound process distribution and functions of this annual loss. The results from this chapter are published in [Targino et al., 2016].

1.3 List of publications and pre-prints

Publications

1. G.W. Peters; **R.S. Targino**; P. Shevchenko *Understanding Operational Risk Capital Approximations: First and Second Orders* (2013) Governance and Regulation, 2(3) (arXiv:1303.2910). (Invited Special Issue to coincide with 8th International conference "International Competition in Banking: Theory and Practice", Sumy, Ukraine, 2013.)

2. **R.S. Targino**; G.W. Peters; P. Shevchenko *Sequential Monte Carlo Samplers for capital allocation under copula-dependent risk models* (2015). Insurance: Mathematics and Economics, 61 (doi:10.1016/j.insmatheco.2015.01.007).

3. **R.S. Targino**; G.W. Peters; G. Sofronov; P. Shevchenko *Optimal exercise strategies for operational risk insurance via multiple optimal stopping times* (2016). Methodology and Computing in Applied Probability (doi:10.1007/s11009-016-9493-8).

Pre-prints

1. G.W. Peters; **R.S. Targino**; M.V. Wüthrich *Full Bayesian Analysis of Claims Reserving Uncertainty* (2016) Available at SSRN 2783223, version of 20/May/2016.

2. G.W. Peters; **R.S. Targino**; M.V. Wüthrich *Bayesian modelling and allocation of insurance risks.* (2016) Manuscript in preparation.

3. C. Chimisov; **R.S. Targino**; G.W. Peters *Risk allocation and risk parity portfolios in high dimensions using Gibbs samplers* (2016) Manuscript in preparation.

Chapter 2

Copulas and risk allocation

This chapter present some background material in the theory of copulas and the mathematical formulation of the capital allocation problem. In particular, we describe the classes of copulas that are used throughout the thesis, as well as some of their properties, with special focus on measures of dependence. In this regard, Section 2.1.3.1 provides a complete picture of the bounds on correlations under a multivariate model described by a Gaussian copula and log-normal marginals. The formulation of the capital allocation problem is provided in Section 2.2. In this section we also discuss how to "coherently" perform the allocation process in a hierarchical structure.

2.1 Copulas and Sklar's theorem

Although the concept of "copula" can be traced back to the seminal work of Abe Sklar [Sklar, 1959], where the mathematical term was introduced, (see [Sklar, 1996]) or even the earlier works of Wassily Hoeffding and Maurice Fréchet, its importance in Finance / Actuarial Science was only realized in the late 1980's / early 1990's. The reader is referred to [Dall'Aglio et al., 1991] for a discussion on the early contributions to the field of copulas; [Frees and Valdez, 1998] and [Embrechts et al., 2002] for some of the publications that boosted the actuarial and financial applications; and [Joe, 1997], [Cherubini et al., 2004], [Nelsen, 2007] and [Joe, 2014] for book-length introductions to the topic.

As with any scientific field, though, copulas were not a unanimity amongst all the researchers and the field saw some interesting academic debate in the past decade, including the one sparked by [Mikosch, 2006] and followed up by academic responses and a rejoinder in the same journal. Unfortunately we also witnessed some shallow non-academic discussion after the 2006+ financial crisis, led by the (in)famous 2009's Wired magazine article [Salmon, 2009] where the author blames the Gaussian copula (see Section 2.1.2 below) model of [Li, 2000] for the financial meltdown in 2006+. Some academic responses to this discussion can be found in [Donnelly and Embrechts, 2010] and, more recently, in Paul Embrechts' interview published as [Durante et al., 2015], while a non-technical defense was provided in The EconomistâĂŹs article [Anonymous, 2009].

At this point we pose the same question asked and promptly answered in [Nelsen, 2007]: *What are copulas? From one point a view, copulas are functions that join or "couple" multivariate distribution functions to their one dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional marginals are uniform on the interval (0,1).* The latter definition is formalized below.

Definition 2.1.1 (Copula)**.** *A d-dimensional copula is a distribution function on* $[0,1]^d$ *with uniform marginal distributions.*

From a modelling point of view, the importance of copulas is summarized by the following elegant and fundamental theorem, which shows that a copula can be extracted from every multivariate distribution function and also that the combination of a copula and univariate distributions leads to a well defined multivariate distribution. For a proof see, for example, [McNeil et al., 2010, Theorem 5.3].

Theorem 2.1.2 (Sklar)**.** *Let* $F_{\boldsymbol{X}}$ *be a joint distributions with marginals* $F_1, \ldots, F_d$ *and denote* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$*. Then there exists a copula* $C : [0,1]^d \to [0,1]$ *such that*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C\big(F_1(x_1), \ldots, F_d(x_d)\big), \quad \forall \boldsymbol{x} = (x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d. \tag{2.1}$$

If the marginals are continuous then C *is unique, and given by*

$$C(u_1, \ldots, u_d) = F_{\boldsymbol{X}}(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)).$$

Conversely, if C *is a copula and* $F_1, \ldots, F_d$ *are univariate distributions, then the distribution* $F_{\boldsymbol{X}}$ *defined in (2.1) is a joint distribution function with marginals* $F_1, \ldots, F_d$*.*

Moreover, if we assume that $F_1, \ldots, F_d$ are differentiable, then the joint density function of $\mathbf{X}$ can be written as

$$f_{\mathbf{X}}(\mathbf{x}) = c\big(F_1(x_1), \ldots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i),$$

where

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}$$

and f_i is the density of X_i.

Another important result in the theory of copulas is the so-called Fréchet-Hoeffding bounds Theorem, stated below.

Theorem 2.1.3 (Fréchet-Hoeffding bounds Theorem)**.** *For any d-dimensional copula* C *the following bounds hold:*

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - 1,\, 0\right\} \leq C(\boldsymbol{u}) \leq \min\{u_1, \ldots, u_d\}.$$

The lower and upper Fréchet-Hoeffding bounds are usually denoted, respectively $W(\mathbf{u})$ and $M(\mathbf{u})$.

In the sequel we branch into two distinct ways of creating copula functions. First, we explicitly define the functional form of the copula and later we extract the copula of known multivariate random variables, generating, respectively, explicit and implicit copulas.

2.1.1 Explicit copulas and Archimedean copulas

Before discussing the class of *Archimedean* copulas we first introduce three *fundamental* copulas: the independence, comonotonicity and countermonotonicity.

Definition 2.1.4. *The d-dimensional independence and comonotonicity copulas are defined, respectively, as*

$$\Pi(\boldsymbol{u}) = \prod_{i=1}^{d} u_i, \quad \textit{and} \quad M(\boldsymbol{u}) = \min\{u_1, \ldots, u_d\}.$$

The countermonotonicity copula is the two-dimensional copula defined as

$$W(\boldsymbol{u}) = \max\{u_1 + u_2 - 1,\, 0\}.$$

It can be seen from Sklar's Theorem that continuous random variables are independent if, and only if, its copula is the independence copula from Definition 2.1.4.

The comonotonicity copula from Definition 2.1.4 is precisely the Fréchet-Hoeffding upper bound, while the countermonotonicity copula is the two-dimensional Fréchet-Hoeffding lower bound (see [McNeil et al., 2010, Example 5.21] for a proof that for $d > 2$ the Fréchet-Hoeffding lower bound is not a copula).

Based on the comonotonic and countermonotonic copulas we now define two important concepts of dependence: comonotonicity and countermonotonicity.

Definition 2.1.5 (Comonotonicity / Countermonotonicity)**.** *The random variables $X_1, \ldots, X_d$ are said to be comonotonic if they admit the Fréchet-Hoeffding upper bound as copula.*

The random variables X_1 and X_2 are said to be countermonotonic if they admit the Fréchet-Hoeffding lower bound as copula.

The following two properties give some insight on the precise meaning of the comonotonicity / countermonotonicity concepts. The proofs can be found in [McNeil et al., 2010, Proposition 5.6 and Proposition 5.19].

Proposition 2.1.6. *The random variables $(X_1, \ldots, X_d)$ are comonotonic if, and only if,*

$$(X_1, \ldots, X_d) \overset{d}{=} (v_1(Z), \ldots, v_d(Z)),$$

for some random variable Z and increasing functions $v_1, \ldots, v_d$.

The random variables X_1 and X_2 are countermonotonic if, and only if,

$$(X_1, X_2) \overset{d}{=} (t_1(Z), t_2(Z)),$$

for some random variable Z with t_1 increasing and t_2 decreasing or vice-versa.

In order to introduce a class of copulas called *Archimedean* copulas we first define the concept of the *generator* of a copula, which is a function of a parameter θ. Some commonly used Archimedean generators are presented in Table 2.1.

Definition 2.1.7 (Archimedean generator)**.** *An Archimedean generator is a continuous, decreasing function* $\psi_\theta : [0, \infty] \to [0, 1]$ *that satisfies* $\psi_\theta(0) = 1$, $lim_{t\to\infty}\psi_\theta(t) = 0$ *and is strictly decreasing on* $[0,\ \inf\{t\ :\ \psi_\theta(t) = 0\}]$.

Definition 2.1.8 (Archimedean copulas)**.** *A d-dimensional copula is called Archimedean if it is of the form*

$$C(\boldsymbol{u};\ \psi_\theta) = \psi_\theta(\psi_\theta^{-1}(u_1) + \ldots + \psi_\theta^{-1}(u_d)), \quad \boldsymbol{u} = (u_1, \ldots, u_d) \in [0, 1]^d, \tag{2.2}$$

where ψ_θ *is the Archimedean generator.*

Remark 2.1.9. *The name* Archimedean *in the context of copulas come, as discussed in [Nelsen, 2007, Section 4.3], from the Archimedean "axiom" from abstract algebra. The "axiom" is indeed a* property *held by some algebraic structures, but the name was coined by the Austrian mathematician Otto Stolz, as it appears in Archimedes' "On the sphere and cylinder" work as Axiom V.*

Family	Parameter	Generator $\psi_\theta(t)$
Clayton	$\theta \in (0, \infty)$	$(1 + t)^{-1/\theta}$
Gumbel	$\theta \in [1, \infty)$	$\exp\{-t^{1/\theta}\}$

Table 2.1: Commonly used Archimedean generators

Although Archimedean copulas may be sufficiently flexible for low (2 to 5) dimensions, it becomes very restrictive as the dimensionality increases, as there is typically a single parameter driving all the dependence structure. Another drawback of Archimedean copulas is the fact that the dependency is symmetric with respect to permutation of variables.

Several alternative classes of copulas have been proposed in the literature lately, including *pair copulas* (see [Aas et al., 2009]), *factor copulas* (see [Oh and Patton, 2013] and [Krupskii and Joe, 2013]) and Hierarchical Archimedean Copulas (HAC), also known as nested Archimedean copulas (see, for example, [Embrechts et al., 2003, Section 6.5] and [Hofert, 2010]). We briefly describe the latter in the sequel, based on [Okhrin and Ristig, 2014].

HAC considers the composition of simple Archimedean copulas (as the ones described in Table 2.1) as follows. A d-dimensional HAC is denoted by $C(u_1, \ldots, u_d;\ s,\ \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the vector of feasible dependency parameters (see discussion below). The parameter $s = (\ldots(i_g i_k) i_l \ldots)$ denotes the structure of the entire HAC, where $i_m \in \{1, \ldots, d\ :\ g \neq k \neq l\}$ is a reordering of the indexes of the variables with $m = 1, \ldots, d$ and $g, k, l \in \{1, \ldots, d\ :\ g \neq k \neq l\}$. Structures of *subcolulas* are denoted by s_j with $s = s_{d-1}$. An example, taken from [Okhrin and Ristig, 2014] is presented in Figure 2.1.

Generators for HAC may come from different families of Archimedean copulas, but care should be taken, as the resulting structure may not be a copula (see [McNeil, 2008, Theorem 4.4]). For generators within the same family, a sufficient condition for the HAC to be a proper

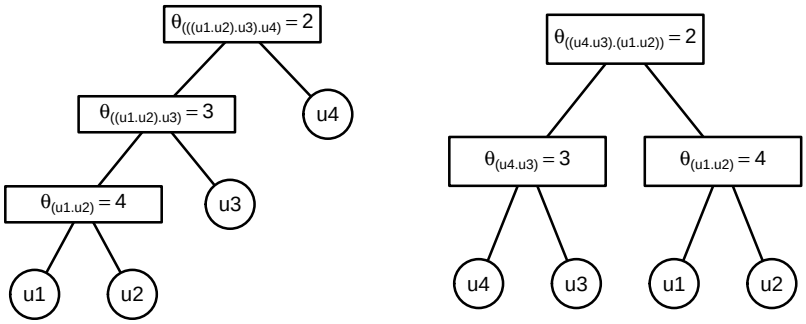

Figure 2.1: Fully (left) and partially (right) nested Archimedean copulas of dimension $d = 4$ with $s = (((12)3)4)$ and $s = ((43)(12))$, respectively. Figure taken from [Okhrin and Ristig, 2014].

copula is to have decreasing parameters from the highest to the lowest level (see, for example, [Hofert, 2010]).

2.1.2 Implicit copulas and the Gaussian copula

In order to characterize the well known Gaussian copula we first discuss the meaning of implicit copulas, extracted from known multivariate distributions.

Definition 2.1.10. *(Copula of F) If $\boldsymbol{Y}$ has joint distribution function $F_{\boldsymbol{Y}}$ with continuous marginal distributions $F_{Y_1}, \dots, F_{Y_d}$, then the copula of F (also called the copula of $\boldsymbol{Y}$) is defined as the distribution function of $\big(F_{Y_1}(Y_1), \dots, F_{Y_d}(Y_d)\big)$, i.e.,*

$$\begin{aligned} C_{\boldsymbol{Y}}(\boldsymbol{u}) &= \mathbb{P}[F_{Y_1}(Y_1) \leq u_1, \dots, F_{Y_d}(Y_d) \leq u_d] \\ &= \mathbb{P}[Y_1 \leq F_{Y_1}^{-1}(u_1), \dots, Y_d \leq F_{Y_d}^{-1}(u_d)] \\ &= F_{\boldsymbol{Y}}\big(F_{Y_1}^{-1}(u_1), \dots, F_{Y_d}^{-1}(u_d)\big). \end{aligned}$$

Given this definition of the implicit copula of F, we can now derive the density of this copula, namely

$$\begin{aligned} c(u_1, \dots, u_d) &= \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d} \\ &= \frac{f_{\mathbf{Y}}\big(F_{Y_1}^{-1}(u_1), \dots, F_{Y_d}^{-1}(u_d)\big)}{f_{Y_1}(F_{Y_1}^{-1}(u_1)) \times \dots \times f_{Y_d}(F_{Y_d}^{-1}(u_d))}. \end{aligned}$$

The next proposition (see [McNeil et al., 2010, Proposition 5.6] for a proof) states an important result related to transformations of marginals.

Proposition 2.1.11 (Invariance)**.** *Let $(Y_1, \dots, Y_d)$ be a random vector with continuous marginals and copula C and let $T_1, \dots, T_d$ be strictly increasing functions. Then $\big(T_1(Y_1), \dots, T_d(Y_d)\big)$ also has copula C.*

Using Sklar's Theorem we can now define the so-called Gaussian copula, which is part of the class of implicit copulas (those extracted from known multivariate distributions).

Definition 2.1.12. *(Gaussian copula) If $\mathbf{Y} \sim N(\mu, \Sigma)$ denotes a multivariate gaussian random variable with mean vector μ and covariance matrix Σ, then its copula is called the Gaussian copula.*

Note that if we define $P = \rho(\Sigma)$ as the correlation matrix of $\mathbf{Y}$ then Proposition 2.1.11 ensures that the copula for $\mathbf{Y}$ is the same as the copula of $\mathbf{X} \sim N(0, P)$. Therefore, from Definition 2.1.10 the Gaussian copula is given by

$$\begin{aligned} C_{\Sigma}^{Ga}(\mathbf{u}) &= \mathbb{P}[\Phi(X_1) \leq u_1, \ldots, \Phi(X_d) \leq u_d] \\ &= \Phi_{0,P}\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big), \end{aligned}$$

with density

$$c_{\Sigma}^{Ga}(\mathbf{u}) = \frac{\phi_{0,P}\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big)}{\phi(\Phi^{-1}(u_1)) \times \cdots \times \phi(\Phi^{-1}(u_d))}$$

where $\Phi_{\mu,\Sigma}(\cdot)$ and $\phi_{\mu,\Sigma}(\cdot)$ denotes, respectively, a (multivariate) normal distribution and density functions with mean μ and covariance matrix Σ. Without explicit mention to μ and Σ, $\Phi(\cdot)$ and $\phi(\cdot)$ denote, respectively, the standard univariate normal distribution and density. If we define $\mathbf{q} = (q_1, \ldots, q_d)$ as the normal scores, i.e., $q_i = \Phi^{-1}(u_i)$ then the density of the Gaussian copula can be simplified to

$$\begin{aligned} c_{\Sigma}^{Ga}(\mathbf{u}) &= \frac{(2\pi)^{-d/2}|P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{q}^T P^{-1}\mathbf{q}\right\}}{\prod_{i=1}^{d}(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}q_i^2\right\}} \\ &= |P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{q}^T(P^{-1} - \mathbf{I})\mathbf{q}\right\}, \end{aligned}$$

where $|P| = det(P)$. Note that the Gaussian copula is parametrized by the $d(d-1)/2$ parameters of the correlation matrix.

In summary, if we want a multivariate random variable $\mathbf{X} = (X_1, \ldots, X_d)$ to have arbitrary marginals $F_{X_1}, \ldots, F_{X_d}$ and the dependence structure to be given by a Gaussian copula with correlation matrix P, then its distribution and density functions should be, respectively,

$$F_{\mathbf{X}}(\mathbf{x}) = \Phi_{0,P}\big(\Phi^{-1}(F_{X_1}(x_1)), \ldots, \Phi^{-1}(F_{X_d}(x_d))\big);$$

$$f_{\mathbf{X}}(\mathbf{x}) = |P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{q}^T(P^{-1} - \mathbf{I})\mathbf{q}\right\}\prod_{i=1}^{d} f_{X_i}(x_i),$$

where here the normal scores are defined as $q_i = \Phi^{-1}\big(F_{X_i}(x_i)\big)$.

Alternatively, let us now assume we have a r.v. $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$, with $(\Sigma)_{i,i} = \sigma_i^2$ and $(\Sigma)_{i,j} = 0$ if $i \neq j$. In other words, $X_i \sim N(0, \sigma_i^2)$ and X_i is independent of X_j $\forall j \neq i$. Given this model for $\mathbf{X}$, we want to find the correlation matrix P of the Gaussian copula that couples the marginals $X_i \sim N(0, \sigma_i^2)$ and return us a joint distribution $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$.

First note that if $X_i \sim N(0, \sigma_i^2)$ then its distribution and density functions can be written in terms of standard normal's c.d.f. and p.d.f. as follows:

$$F_i(x_i) = \Phi(x_i/\sigma_i), \qquad f_i(x_i) = \phi(x_i/\sigma_i)\sigma_i^{-1}.$$

Therefore the joint p.d.f. of $\mathbf{X}$ can be written as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= c_P^{Ga}\big(F_1(x_1),\ldots,F_d(x_d)\big)\prod_{i=1}^{d} f_i(x_i) \\ &= |P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{q}^T(P^{-1}-\mathbf{I})\mathbf{q}\right\}\prod_{i=1}^{d}\phi(x_i/\sigma_i)\sigma_i^{-1}, \end{aligned}$$

where $q_i = \Phi^{-1}\big(F_i(x_i)\big) = \Phi^{-1}\big(\Phi(x_i/\sigma_i)\big) = x_i/\sigma_i$.

From the above formula, however, it is not clear what should be the choice for P in order to have $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$. To overcome this difficulty, let $\mathbf{W}$ be such that $W_i = \frac{X_i}{\sigma_i}$, for $i = 1, \ldots, d$. It then implies that $X_i = \sigma_i W_i$ and that

$$\left|\frac{\partial \mathbf{x}}{\partial \mathbf{w}}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial w_1} & \cdots & \frac{\partial x_1}{\partial w_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial w_1} & \cdots & \frac{\partial x_d}{\partial w_d} \end{vmatrix} = \begin{vmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_d \end{vmatrix} = \prod_{i=1}^{d}\sigma_i$$

Therefore, the p.d.f. of the transformed random variable is given by

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{w}) &= f_{\mathbf{X}}(\sigma_1 w_1,\ldots,\sigma_d w_d)\prod_{i=1}^{d}\sigma_i \\ &= |P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{w}^T P^{-1}\mathbf{w}\right\}\exp\left\{-\frac{1}{2}\mathbf{w}^T\mathbf{w}\right\}\prod_{i=1}^{d}\phi(w_i) \\ &= (2\pi)^{-d/2}|P|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{w}^T P^{-1}\mathbf{w}\right\}. \end{aligned}$$

From this relationship we can deduce that $\mathbf{W} \sim N(\mathbf{0}, P)$ thus we can see that $\mathbf{X} \sim N(0, \Sigma)$ with

$$\Sigma_{i,i} = Var(\sigma_i W_i) = \sigma_i^2 Var(W_i), \qquad \Sigma_{i,j} = Cov(X_i, X_j) = \sigma_i\sigma_j Cov(W_i, W_j).$$

Since the variance-covariance matrix of $\mathbf{W}$, namely P, is a correlation matrix we have that $P_{i,i} = 1$ and $P_{i,j} = \Sigma_{i,j}/(\sigma_i\sigma_j)$.

2.1.3 Dependence measures

As we saw in Sklar's Theorem, the copula function encompasses the whole dependence structure of a multivariate random variable. Therefore, it is natural to develop scalar dependence measures that explain, in some sense, the strength of the dependence. In the sequel we discuss three different types of dependence measures: (1) linear correlation; (2) rank correlation; and (3) coefficients of tail dependence (and its extension to the multivariate case).

2.1.3.1 Linear correlation

Linear correlation is certainly one of the most well known concepts in Statistics and some misconceptions around it are discussed in length in [Embrechts et al., 2002]. In this section we focus on one particular point, related to the attainability of prescribed correlations on a specific model, as it is an important regulatory requirement (as discussed in Chapter 6).

As discussed, for example in [Embrechts et al., 2002, Fallacy 2], for given marginal distributions not all linear correlations between -1 and 1 can be achieved. This can also be seen in the following Lemma (see [Denuit and Dhaene, 2003, Section 2]).

Lemma 2.1.13 (Correlation bounds)**.** *Let (X_1, X_2) be a bivariate random variable with marginal distributions F_1 and F_2. Then the correlation between X_1 and X_2 is bounded by*

$$\frac{Cov(F_1^{-1}(U),\ F_2^{-1}(1-U))}{\sqrt{Var(X_1)\,Var(X_2)}} \leq Corr(X_1,\ X_2) \leq \frac{Cov(F_1^{-1}(U),\ F_2^{-1}(U))}{\sqrt{Var(X_1)\,Var(X_2)}}$$

for U uniformly distributed in $[0,1]$.

Although theoretically interesting, Lemma 2.1.13 may provide bounds that are too wide and in some cases just state that the correlation lies between -1 and 1. In the sequel we show that in the particular case of a random vector with Log-Normal marginals and dependence structure Gaussian copula it is possible to calculate precisely the intended correlation and numerically check its limits.

Let us assume a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ is normally distributed with $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{m},\ \boldsymbol{V})$, where a general term of the covariance matrix $\boldsymbol{V}$ is given by $(\boldsymbol{V})_{i,j} = V_{i,j}$ and $V_{i,i} = V_i^2$. Moreover, we denote by $\boldsymbol{\Omega} = \mathrm{Corr}(\boldsymbol{X})$ the correlation matrix of the random vector $\boldsymbol{X}$, i.e.,

$$\boldsymbol{V} = \mathrm{diag}(V_1, \ldots, V_d)\ \boldsymbol{\Omega}\ \mathrm{diag}(V_1, \ldots, V_d),$$

with $(\boldsymbol{\Omega})_{i,j} = (\boldsymbol{\Omega})_{j,i} = \omega_{i,j}$.

If we define $Z_i = e^{X_i}$, for $i = 1, \ldots, d$ then $Z_i \sim \mathrm{LN}(m_i,\ V_i)$ with

$$\begin{aligned} \mathbb{E}[Z_i] &= \exp\left\{m_i + \frac{V_i^2}{2}\right\} \\ \mathrm{Var}(Z_i) &= (\mathbb{E}[Z_i])^2 \left(e^{V_i^2} - 1\right). \end{aligned} \tag{2.3}$$

On the other hand, since $X_i + X_j \sim N(m_i + m_j,\ V_i^2 + V_j^2 + 2V_i\omega_{i,j}V_j)$ we have that

$$\mathbb{E}[Z_i Z_j] = \mathbb{E}[e^{X_i + X_j}] = \exp\left\{m_i + m_j + \frac{V_i^2 + V_j^2 + 2V_i\omega_{i,j}V_j}{2}\right\}. \tag{2.4}$$

Therefore, using (2.3) and (2.4) the correlation between Z_i and Z_j can be written as

$$\mathrm{Corr}(Z_i,\ Z_j) = \frac{\exp\{V_i\omega_{i,j}V_j\} - 1}{\left[(e^{V_i^2} - 1)(e^{V_j^2} - 1)\right]^{1/2}}. \tag{2.5}$$

Since $\exp(\cdot)$ is a strictly increasing function and the marginal distributions of $(X_1, \ldots, X_d)$ are continuous, from Proposition 2.1.11 we can conclude that $(Z_1, \ldots, Z_d)$ has the same copula as $(X_1, \ldots, X_d)$: a Gaussian copula with correlation matrix $\boldsymbol{\Omega}$.

From equation (2.5) it is easy to see the correlation between Z_i and Z_j is a monotone function of $\omega_{i,j}$ which implies that $\mathrm{Corr}(Z_i,\ Z_j)$ will be minimal when $\omega_{i,j} = -1$ and maximal when $\omega_{i,j} = 1$. Hence, for a given pair of standard deviations it is possible to compute the interval of admissible correlations for the pair $(Z_i,\ Z_j)$. On Figure 2.2 the lower (left plot) and upper (right plot) bounds for the correlations are presented.

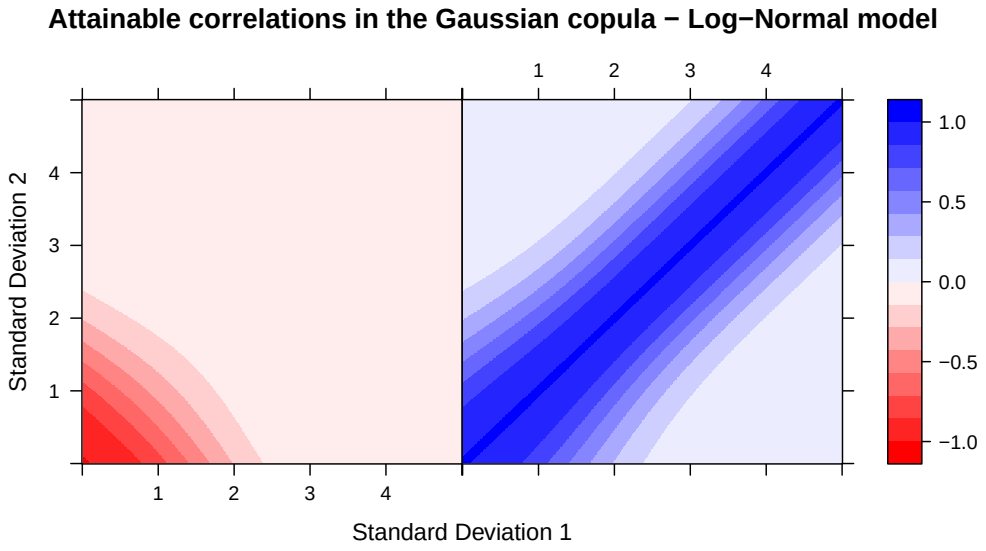

Figure 2.2: Lower (left) and upper (right) bound for correlations in a Gaussian-copula model with Log-Normal marginal distributions, as a function of the scale parameters σ_1 and σ_2.

Figure 2.2 shows that even when the copula correlation is set to -1, if at least one of the standard deviation parameters is "large", then the minimum possible correlation between the log-normal variables is close to zero. For example, if $\sigma_1 = \sigma_2 = 2$ then the lower bound for the correlation between these variables is approximately -2%. As actuarial risks are usually positively correlated this may not be a problem from the modelling point of view. In contrast to the lower limit, the upper limit for the correlations have a different behaviour. If both standard deviations are the same, then the range of attainable correlations is upper bounded by 1, meaning that any positive correlation can be achieved. Problems arise when the standard deviations are sufficiently different from each other. If $\sigma_1 = 1$, then the correlation is upper bounded by 66% if $\sigma_2 = 2$, 16% if $\sigma_2 = 3$ and about 1% if $\sigma_2 = 4$. From the actuarial example discussed in Chapters 5 and 6 (see Table 6.5) we have that the largest difference between standard deviations is for $\sigma_1 \approx 0.020$ and $\sigma_2 \approx 0.20$ and, in this case, the range of attainable correlations is given by $[-0.997, 0.991]$ which safely spans all practically relevant values.

In [Devroye and Letac, 2015] the authors discuss a similar problem. Let us denote by $\mathcal{R}_n$ the set of all $n \times n$, symmetric, positive semi-definite matrices with diagonal terms equal to 1; and by $R(C) = \text{Corr}(\mathbf{U})$ the correlation matrix of a random vector $\mathbf{U} \sim C$, with elements $U_i \sim [0, 1]$. The question asked in [Devroye and Letac, 2015] is: given $R \in \mathcal{R}_n$, does there exist a copula C such that $R(C) = R$? The answer is *yes*, if $n \leq 9$ and the authors postulate that for $n \geq 10$ there exists $R \in \mathcal{R}_n$ such that there is no copula C such that $R(C) = R$.

2.1.3.2 Rank correlation

As the linear correlation coefficient, the rank correlation is also a scalar measure of dependence of two random variables. The main difference between these two measures is that the quantity

being introduced in this section only depends on the copula, instead of the joint distribution (i.e., the copula and the marginals).

Before discussing the two main forms of rank correlation, namely Kendall's τ and Spearman's ρ, we notice that rank correlations, as the name implies, are statistics that depend only on the *ranks* (i.e., the order) of the data and not on its specific values.

Definition 2.1.14 (Concordant/discordant points)**.** *Two points* $\boldsymbol{x} = (x_1, x_2)$, $\widetilde{\boldsymbol{x}} = (\widetilde{x}_1, \widetilde{x}_t) \in \mathbb{R}^2$ *are said to be concordant if* $(x_1 - \widetilde{x}_1)(x_2 - \widetilde{x}_2) > 0$ *and discordant if* $(x_1 - \widetilde{x}_1)(x_2 - \widetilde{x}_2) < 0$.

If $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2)$ is an independent copy of $\mathbf{X} = (X_1, X_2)$ then the Kendall's τ is defined as the difference between the probability of concordance and the probability of discordance, which can also be written as an expectation.

Definition 2.1.15 (Kendall's τ)**.** *The Kendall's* τ *of a bivariate random variable* (X_1, X_2) *is defined as*

$$\begin{aligned}\tau(\boldsymbol{X}) &= \mathbb{P}[(X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0] - \mathbb{P}[(X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0] \\ &= \mathbb{E}[sign\big((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2)\big)],\end{aligned}$$

where $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ *is an independent copy of* $\boldsymbol{X}$.

More generally, for $\boldsymbol{X} \in \mathbb{R}^d$,

$$\tau(\boldsymbol{X}) = Cov(sign(\boldsymbol{X} - \widetilde{\boldsymbol{X}})).$$

For the Spearman's ρ we follow [McNeil et al., 2010, Definition 5.28], where it is defined as the linear correlation coefficient of the probability-transformed random variables.

Definition 2.1.16 (Spearman's ρ)**.** *For continuous random variables* X_1 *and* X_2 *with marginal cdfs* F_{X_1} *and* F_{X_2} *the Spearman's* ρ *is given by*

$$\rho(\boldsymbol{X}) = Corr(F_{X_1}(X_1),\, F_{X_2}(X_2)).$$

From the discussion above it can be seen that all three quantities presented (linear correlation, Kendall's τ and Spearman's ρ) are symmetric, defined in the range $[-1, 1]$ and give value of zero for independent random variables. Also, as in the well known case of linear correlation, a rank correlation of zero does not imply independence. Moreover, as shown in [Embrechts et al., 2002] the rank correlation take the value of 1 when X_1 and X_2 are comonotonic and -1 when they are countermonotonic.

As previously stated, the next result shows that both the Kendall's τ and the Spearman's ρ only depend on the copula function. For a proof see [McNeil et al., 2010, Proposition 5.29].

Proposition 2.1.17. *Assume* X_1 *and* X_2 *have continuous marginal distributions. Then the rank correlations are given by*

$$\begin{aligned}\tau(X_1, X_2) &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1, \\ \rho(X_1, X_2) &= 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2.\end{aligned}$$

An interesting on the comparison of Spearman's ρ and Kendall's τ is found in [Capéraà and Genest, 1993], where the authors prove that Spearman's ρ is larger than Kendall's τ whenever X_2 is right-tail increasing in X_1 and X_2 is left-tail decreasing in X_1 (or reversed roles of X_1, X_2), where X_2 is right-tail increasing in X_1 if $\mathbb{P}[X_2 > x - 2|X_1 > x_1]$ is nondecreasing in x_1 for all x_2. The concept of left-tail decreasing is defined analogously. Other comparative results have been proved in the literature, see, for example [Fredricks and Nelsen, 2007] in references therein. Apart from proving that $\rho(X_1, X_2)/\tau(X_1, X_2)$ converges to 3/2 as the joint distribution approaches that of two independent random variables, a simpler proof of the result in [Capéraà and Genest, 1993] is also given.

2.1.3.3 Coefficient of tail dependence

As the rank correlation, the coefficients of tail dependence are measures of dependence that depend only on the copula of a pair of random variables and not on its marginals.

Informally, the tail dependence is a measure of strength of the dependence in the joint lower or upper tails of a multivariate distribution. The upper tail dependence is defined as the limit of the probability of a random variable X_2 exceeds its α-quantile given that X_1 exceeds its own α-quantile, where the limit is taken when α grows towards 1. A similar definition is also provided for the lower tail dependence, as seen below.

Definition 2.1.18 (Coefficients of tail dependence)**.** *For X_1 and X_2 continuous random variables with distributions F_1 and F_2 the coefficients of upper and lower tail dependence are defined, respectively, as*

$$\lambda_u(X_1, X_2) = \lim_{\alpha \to 1^-} \mathbb{P}[X_1 > F_1^{-1}(\alpha) \,|\, X_2 > F_2^{-1}(\alpha)]$$
$$\lambda_l(X_1, X_2) = \lim_{\alpha \to 0^+} \mathbb{P}[X_1 \leq F_1^{-1}(\alpha) \,|\, X_2 \leq F_2^{-1}(\alpha)],$$

provided these limits exist.

When $\lambda_u(X_1, X_2) \in (0, 1]$ we say the two variables have upper tail dependence and when $\lambda_u(X_1, X_2) = 0$ we say they are asymptotically independent in the upper tail. Analogous nomenclatures are used for the lower tail dependence.

Rearranging the coefficients in terms of the copula of (X_1, X_2) we see that

$$\lambda_u(X_1, X_2) = \lim_{\alpha \to 1^-} \frac{\mathbb{P}[X_1 > F_1^{-1}(\alpha),\, X_2 > F_2^{-1}(\alpha)]}{\mathbb{P}[X_2 > F_2^{-1}(\alpha)]} = \lim_{\alpha \to 1^-} \frac{C(1-\alpha,\, 1-\alpha)}{\alpha}$$
$$\lambda_l(X_1, X_2) = \lim_{\alpha \to 0^+} \frac{\mathbb{P}[X_1 \leq F_1^{-1}(\alpha),\, X_2 \leq F_2^{-1}(\alpha)]}{\mathbb{P}[X_2 \leq F_2^{-1}(\alpha)]} = \lim_{\alpha \to 0^+} \frac{C(\alpha,\, \alpha)}{\alpha}.$$

As an extension of these concepts we follow [De Luca and Rivieccio, 2012] and the define the multivariate coefficients of tail dependence as follows.

Definition 2.1.19 (Multivariate coefficients of tail dependence)**.** *For $X_1, \ldots, X_d$ continuous random variables with distributions $F_1, \ldots, F_d$ the multivariate coefficients of upper and lower*

tail dependence are defined, respectively, as

$$\lambda_u = \lambda_u(X_1,\dots,X_d) = \lim_{\alpha\to 1^-} \mathbb{P}[X_1 > F_1^{-1}(\alpha) \,|\, X_2 > F_2^{-1}(\alpha),\dots,X_d > F_d^{-1}(\alpha)]$$
$$\lambda_l = \lambda_l(X_1,\dots,X_d) = \lim_{\alpha\to 0^+} \mathbb{P}[X_1 \le F_1^{-1}(\alpha) \,|\, X_2 \le F_2^{-1}(\alpha),\dots,X_d \le F_d^{-1}(\alpha)],$$

provided these limits exist.

For the Archimedean copulas defined in Table 2.1 the multivariate coefficients of upper and lower tail dependence were shown in [De Luca and Rivieccio, 2012] to be as follows.

Proposition 2.1.20. *For the Archimedean copulas defined in Table 2.1 the coefficients of multivariate tail dependence are given by*

$$\lambda_u = \lim_{t\to 0^+} \frac{\sum_{i=1}^{d} \binom{n}{n-i} i(-1)^i \psi'(it)}{\sum_{i=1}^{n-1} \binom{n-1}{n-1-i} i(-1)^i \psi'(it)}$$
$$\lambda_l = \frac{d}{d-1} \lim_{t\to\infty} \frac{\psi'(dt)}{\psi'\big((d-1)t\big)}.$$

For completeness in Table 2.2 we present the dependence measures which are only a function of the copula, for the Gaussian, Gumbel and Clayton copulas. Dashed cells denote dependence measures which are not known in closed form (see [Kruskal, 1958] and [McNeil et al., 2010, Table 5.5]).

Copula	$\rho(X_1,X_2)$	$\tau(X_1,X_2)$	$\lambda_l(X_1,X_2)$	$\lambda_u(X_1,X_2)$
Gaussian	$\frac{6}{\pi} arcsin\left(\frac{\mathrm{Corr}(X_1,X_2)}{2}\right)$	$\frac{2}{\pi} arcsin\big(\mathrm{Corr}(X_1,X_2)\big)$	0	0
Gumbel	——	$1-1/\theta$	0	$2-2^{1/\theta}$
Clayton	——	$\theta/(\theta+2)$	$2^{-1/\theta},\ \theta>0$ $0,\ \theta\le 0$	0

Table 2.2: Dependence measures (Spearman's ρ, Kendall's τ, lower and upper tail dependence) for Gaussian, Gumbel and Clayton copulas.

2.2 Risk contributions and capital allocations

In this section we characterize mathematically some of the concepts discussed in Chapter 1. In particular, we discuss the concept of *risk* reviewing the concept of *coherent risk measures* and discuss how to *allocate* risks to different components of a portfolio.

First, let us assume $X_1,\dots,X_d$ are positive random variables representing the aggregated losses, in a fixed time horizon, of d different assets in a portfolio. In the case of OpRisk modelling (see Chapter 4), this notation corresponds to operational losses from d different business unit and risk type combinations within divisions of a banking or insurance institution. For the actuarial application of Chapter 6 these variables denote claims payment for different lines of business. In both cases the time horizon is defined as one year, all the random variables are

defined in a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{X}$ is assumed to be the set of all possible financial positions (i.e., random variables such as X_i).

If the weights of these positions in a portfolio are given by $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \Lambda \subset \mathbb{R}^d \setminus \{\mathbf{0}\}$, then we denote the portfolio-wide loss by

$$S(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i X_i. \tag{2.6}$$

In particular, if $\boldsymbol{\lambda} = (1, \ldots, 1)$ we write $S = S(\boldsymbol{\lambda})$. In the case of OpRisk modelling this aggregate loss amount given by $\sum_{i=1}^{d} \lambda_i X_i$ represents the institution-wide total annual loss and typically the weights would be equal to one.

In this context a *risk measure* is a functional $\rho : \mathcal{X} \to \mathbb{R}$, which assigns a real value for each financial position X. Given a specific risk measure ρ, the function $r_\rho : \Lambda \to \mathbb{R}$ such that $r_\rho(\boldsymbol{\lambda}) = \rho(S(\boldsymbol{\lambda}))$ is called a *risk-measure function*. Three of the most popular risk measures amongst practiotioners, regulators and academics are given in Definition 2.2.1.

Definition 2.2.1 (Particular choices of risk measures)**.** *If $X_1, \ldots, X_d$ are continuous random variables, and $S = \sum_{i=1}^{d} X_i$ is also continuous, three of the most popular choices of risk measures are given by*

1. *Standard deviation: $\rho(S) = \sqrt{Var(S)}$;*
2. *Value at Risk: $\rho(S) = VaR_\alpha(S) := \inf\{s \in \mathbb{R} : F_S(s) = \alpha\}$;*
3. *Expected Shortfall: $\rho(S) = ES_\alpha(S) := \mathbb{E}[S \,|\, S \geq VaR_\alpha(S)]$.*

Although reasonable at a first glance, the risk measures from Definition 2.2.1 were proposed in the financial risk management community in a rather ad-hoc manner. In the seminal paper [Artzner et al., 1999] the authors take a step back and discuss four properties a "good" risk measure should have: monotonicity, translation invariance, subadditivity and positive homogeneity. Before presenting these properties we recall the definition of a *homogeneous function*, closely related to the positive homogeneity of [Artzner et al., 1999].

Definition 2.2.2 (Homogeneous function)**.** *A function $f : U \subset \mathbb{R}^d \to \mathbb{R}$ is said to be homogeneous of degree τ if, for all $h > 0$ and $u \in U$ with $hu \in U$ the following equation holds:*

$$f(hu) = h^\tau f(u).$$

For any two random variables X and Y defined in $\mathcal{X}$, the four properties from [Artzner et al., 1999] are described below.

1. Monotonicity: If $X \geq Y$ $\mathbb{P}$–almost surely, then $\rho(X) \geq \rho(Y)$;
2. Translation invariance: If $m \in \mathbb{R}$, then $\rho(X - m) = \rho(X) - m$;
3. Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$;

4. Positive homogeneity: If $\lambda \geq 0$, then $\rho(\lambda X) = \lambda\rho(X)$.

These properties can be interpreted as follows. **Monotonicity** implies that if a position leads (almost surely) larger losses than another one, then it should be considered riskier. **Translation invariance** requires that adding (subtracting) a deterministic amount of capital to a position should increase (decrease) the risk measure by the same amount. **Subadditivity**, which is perhaps the most debatable axiom, requires that the addition of two positions does not create extra risk, reflecting diversification benefits. **Positive homogeneity** requires the risk of a position to increase linearly with its size. Note that it provides, as a consequence, a *normalization* property: $\rho(0) = 0$.

In [Artzner et al., 1999] the authors take the four properties above as axioms to define the coherence of a risk measure, as formalized below.

Definition 2.2.3 (Coherent risk measure)**.** *A risk measure ρ satisfying the* monotonicity, translation invariance, subadditivity and positive homogeneity *properties is called a* coherent risk measure.

For the risk measures from Definition 2.2.1 only the Expected Shortfall is guaranteed to be coherent, since the standard deviation can break the monotonicity axiom and the Value at Risk the subadditivity, as we see below.

A simple example of the lack of monotonicity of the standard deviation is as follows. Let $X, Y \in \mathcal{X}$ be positive random variables such that their standard deviations are, respectively, $\sigma_X = 1$ and $\sigma_Y = 1$ and their covariance (which is also their correlation) is $\sigma_{X,Y} = -0.75$. For $Z = X + Y$ we have that $\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}}$. Therefore, in this example $Z \geq X$ but $\sigma_Z = \sqrt{1 + 1 - 2 \times 0.75} = \sqrt{0.5} < 1 = \sigma_X$.

Several authors discussed the lack of subadditivity of the VaR, including [Artzner et al., 1999]. Next we present an example from [Daníelsson et al., 2013] of the violation of the subadditivity property. For $i \in \{1, 2\}$, let $X_i = \varepsilon_i + \eta_i$, with

$$\varepsilon_i \sim N(0,1) \quad \text{and} \quad \eta_i = \begin{cases} 0, \text{ with probability } 0.991, \\ 10, \text{ with probability } 0.009 \end{cases}$$

Assuming that all combinations of ε_i and η_j are independent for $i, j \in \{1, 2\}$ we have that

$$\begin{aligned} 0.99 &= \mathbb{P}[X_1 \leq \text{VaR}_{99\%}(X_1)] \\ &= \mathbb{P}[\varepsilon_1 \leq \text{VaR}_{99\%}(X_1) - \eta_1] \\ &= \mathbb{P}[\varepsilon_1 \leq \text{VaR}_{99\%}(X_1)] \times 0.991 + \mathbb{P}[\varepsilon_1 \leq \text{VaR}_{99\%}(X_1) - 10] \times 0.009. \end{aligned} \tag{2.7}$$

Denoting $E = \varepsilon_1 + \varepsilon_2$ and $N = \eta_1 + \eta_2$, we have that

$$E \sim N(0, \sqrt{2}) \quad \text{and} \quad N = \begin{cases} 0, \text{ with probability } 0.991^2, \\ 10, \text{ with probability } 2 \times 0.009 \times 0.991, \\ 20, \text{ with probability } 0.009^2 \end{cases}$$

Then,

$$\begin{aligned} 0.99 &= \mathbb{P}[X_1 + X_2 \leq \mathrm{VaR}_{99\%}(X_1 + X_2)] \\ &= \mathbb{P}[E \leq \mathrm{VaR}_{99\%}(X_1 + X_2) - N] \\ &= \mathbb{P}[E \leq \mathrm{VaR}_{99\%}(X_1 + X_2)] \times 0.991^2 \\ &\quad + \mathbb{P}[E \leq \mathrm{VaR}_{99\%}(X_1 + X_2) - 10] \times 2 \times 0.991 \times 0.009 \\ &\quad + \mathbb{P}[E \leq \mathrm{VaR}_{99\%}(X_1 + X_2) - 20] \times 0.009^2. \end{aligned} \tag{2.8}$$

Numerically solving (2.7) for $\mathrm{VaR}_{99\%}(X_1)$ and (2.8) for $\mathrm{VaR}_{99\%}(X_1 + X_2)$ we find that

$$\mathrm{VaR}_{99\%}(X_1 + X_2) = 9.80 > 6.17 = \mathrm{VaR}_{99\%}(X_1) + \mathrm{VaR}_{99\%}(X_2),$$

which shows the lack of subadditivity of the VaR in this example.

For different proofs of the subadditivity of the ES the reader is referred to [Embrechts et al., 2015]. It also worth mentioning the Expected Shortfall has been proposed independently by several authors, under different names, such as Tail VaR (TVaR), Tail Conditional Expectation (TCE) and Conditional VaR (CVaR). For an overview of these definitions we refer the reader to [Acerbi and Tasche, 2002].

Discussions on the usage of the VaR or ES as a risk measure date as far back as the introduction of the latter. At the moment Solvency II requires the Solvency Capital Requirement (SCR) to be calculated based on the one year 99.5% VaR (see [EIOPS, 2010, SCR.1.9]); the SCR in the Swiss Solvency Test (SST) is based on the one year 99% Expected Shortfall (see [FINMA, 2007, Section 2]); for operational risk capital Basel II calculates the one year 99.9% VaR (see [BCBS, 2006, § 667]), but it is on the process of revising the operational risk framework (see [BCBS, 2016b] and Section 1.1.1.3); and for market risk the new Basel requirements have just been agreed to change from 99% VaR to 97.5% ES (see [BCBS, 2016a, § 181]).

The battle between VaR and ES became even more pronounced when [Gneiting, 2011] revealed an issue with direct backtesting of ES estimates, as this risk measure is not *elicitable* (while the VaR is). As a response [Acerbi and Székely, 2014] and [Emmer et al., 2015] discussed different ways to backtest ES, and the latter also discusses the elicitability and other desirable properties (such as the coherence axioms) of VaR, ES and *expectiles*.

On a different direction, [Cont et al., 2010] the authors proposed a *robust* risk measure, the so-called Range-Value-at-Risk (RVaR), defined as

$$\mathrm{RVaR}_{\alpha,\beta}(S) = \begin{cases} \frac{1}{1-\beta}\int_{\alpha+\beta-1}^{\alpha} \mathrm{VaR}_\gamma(S)d\gamma, & \text{if } \beta < 1 \\ \mathrm{VaR}_\alpha(S), & \text{if } \beta = 1 \end{cases},$$

which can be seen as an interpolation between the VaR (for $\beta = 1$) and ES (for $\alpha = 1$). In fact, in [Cont et al., 2010] it has been proved that for $\alpha < 1$ and $\alpha + \beta < 1$, $\mathrm{RVaR}_{\alpha,\beta}$ is continuous with respect to the convergence in distribution.

Several other classes of risk measures have been proposed in the literature, such as *distortion risk measures* ([Denneberg, 1994] and [Wang et al., 1997]); *spectral risk measures*, which is

the sub-class of distortion risk measures which are also coherent ([Acerbi, 2002]); and *convex risk measures* ([Föllmer and Schied, 2002] and [Frittelli and Gianin, 2002]). An introduction to these measures can be found, for example, in [Bignozzi, 2012, Chapter 1].

For the objectives of this work, we step aside from the discussion about different risk measures and mostly focus on ES. If we now assume the risk measure ρ has been chosen and that $\rho(S(\boldsymbol{\lambda}))$, the risk (capital) of the portfolio (institution) has already been calculated we now focus our attention to the process of capital (risk) allocation. But before we state the theorem that gives name to the allocation principle to be discussed.

Theorem 2.2.4 (Euler's homogeneous function theorem)**.** *Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ be a continuously differentiable function. Then f is homogeneous of degree τ if, and only if, it satisfies the following equation:*

$$\tau f(\boldsymbol{u}) = \sum_{i=1}^{n} u_i \frac{\partial f}{\partial u_i}, \quad \boldsymbol{u} = (u_1, \ldots, u_n) \in U,\, h > 0.$$

The allocation process consists of understanding how much of the risk (capital) is due to each of the constituents of a portfolio. In the case of OpRisk, this involves understanding what each division's total capital requirement would be across all the Basel III risk types (see Table 1.4), based on a "fair' allocation of the institution's total capital requirement. For the actuarial application discussed in Chapter 6 our interest is on allocating the capital to lines of business and "old" and "new" risks (in a concept to be make precise in Chapter 5).

One of the reasons for performing the allocation exercise (which is not required by regulators) is to utilize the results for a risk-reward management tool. The amount of capital (or risk) allocated to each line of business, for example, may assist the central management's decision to further invest in or discontinue a business line and even to remunerate line managers. We return to the topic of performance measurement in Section 2.2.3, where we discuss the concept of Return On Risk-Adjusted Capital (RORAC).

In contrast to quantitative risk (capital) assessment, where there is an unanimous view (shared by regulators world-wide, as seen in Chapter 1) that it should be performed through the use of risk measures, such as the VaR or ES, there is no consensus on how to perform capital allocation. In this work we follow the so-called Euler allocation principle (see, e.g., [Tasche, 1999] and [McNeil et al., 2010, Section 6.3]), which is discussed below. For different allocation principles the reader is referred to [Dhaene et al., 2012].

To introduce the allocation principle used throughout this thesis, let us denote by $\mathcal{A}_i^{\rho}(\boldsymbol{\lambda})$ the capital allocated to one unit of X_i when the portfolio's loss is given by $S(\boldsymbol{\lambda})$. For the sake of simplicity, to derive the Euler allocation we accept the following set of assumptions.

Assumptions 2.2.5. *If the individual and portfolio losses are given, respectively by $X_1, \ldots, X_d$ and Equation (2.6), then we assume that*

(i) the capital allocated to the position $\lambda_i X_i$ is given by $\lambda_i \mathcal{A}_i^{\rho}(\boldsymbol{\lambda})$;

(ii) the overall risk capital $r_\rho(\boldsymbol{\lambda})$ is fully allocated to the individual positions in the portfolio:

$$\sum_{i=1}^{d} \lambda_i \mathcal{A}_i^\rho(\boldsymbol{\lambda}) = r_\rho(\boldsymbol{\lambda}). \tag{2.9}$$

The interpretation of the first assumption is that we require proportional positions to have proportional capital shares and the second one ensures the total capital is allocated to the individual positions. As a result of these assumptions, we only need to calculate $\mathcal{A}_i^\rho(\boldsymbol{\lambda})$, for $i = 1, \dots, d$.

Following the nomenclature in [McNeil et al., 2010], we say $\mathcal{A}_i^\rho : \Lambda \to \mathbb{R}^d$ is a per-unit capital allocation principle if (2.9) is satisfied for all $\boldsymbol{\lambda} \in \Lambda$.

So far we have not imposed or assumed any specific requirement on the risk measure used as a base for the allocation principle, but to introduce the popular Euler allocation principle we need to restrict ourselves to the class of positive homogeneous risk measures. Note that even though the Value at Risk is not coherent in the sense of [Artzner et al., 1999] it is positive homogeneous.

For positive homogeneous risk measures it is trivial to show that the associated risk measure function r_ρ satisfies $r_\rho(t\boldsymbol{\lambda}) = t r_\rho(\boldsymbol{\lambda})$. Therefore, applying Euler's homogeneous function theorem (see Theorem 2.2.4) on r_ρ, we have that

$$r_\rho(\boldsymbol{\lambda}) = \sum_{i=1}^{d} \lambda_i \frac{\partial r_\rho}{\partial \lambda_i}(\boldsymbol{\lambda}) \tag{2.10}$$

The combination of (2.9) and (2.10) leads to the so-called Euler allocation principle (sometimes referred as allocation by the gradient), where the capital allocated to the i-th component of the portfolio is given by the partial derivatives, with $\mathcal{A}_i^\rho(\boldsymbol{\lambda}) := \frac{\partial r_\rho}{\partial \lambda_i}(\boldsymbol{\lambda})$.

The Euler allocation principle arises in different contexts in the literature. For example, in [Denault, 2001] and [Kalkbrener, 2005] the Euler principle is motivated by two (different) sets of axioms, leading to coherent allocation principles (for a relationship between coherent risk measures and coherent capital allocations see [Buch and Dorfleitner, 2008]).

Assuming that $X_1, \dots, X_d$ are continuous random variables at the point at which the risk measure is evaluated, we now present some explicit forms of the Euler contributions, based on the different risk measures presented in Definition 2.2.1. A proof based on the original arguments of [Tasche, 1999] can be found in [McNeil et al., 2010, Section 6.3] while [Emmer et al., 2015] suggests a new approach, based on [Delbaen, 2000], which can also derive *expectiles* allocations.

Proposition 2.2.6. *If $S = \sum_{i=1}^{d} X_i$ and $\boldsymbol{X} = (X_1, \dots, X_d)$ has a joint continuous density, then the Euler allocation takes the following form*

1. *Standard deviation:* $\rho(S) = \sqrt{Var(S)} \Longrightarrow \mathcal{A}_i^\sigma(X_i) = \dfrac{Cov(X_i, S)}{\sqrt{Var(S)}}$;
2. *Value at Risk:* $\rho(S) = VaR_\alpha(S) \Longrightarrow \mathcal{A}_i^{VaR}(X_i) = \mathbb{E}[X_i \,|\, S = VaR_\alpha(S)]$;
3. *Expected Shortfall:* $\rho(S) = ES_\alpha(S) \Longrightarrow \mathcal{A}_i^{ES}(X_i) = \mathbb{E}[X_i \,|\, S \geq VaR_\alpha(S)]$.

Remark 2.2.7. *Proposition 2.2.6 is still valid even if the distribution of* $\boldsymbol{X}$ *is not continuous but other technical conditions should be satisfied (see [Tasche, 1999] and [Gouriéroux et al., 2000]).*

2.2.1 Euler allocation for multivariate normal risks

If we assume the vector of risks follows a multivariate normal distribution, i.e., $\mathbf{X} = (X_1, \ldots, X_d) \sim N(\mathbf{0}, \Sigma)$ then, from Definition 2.2.1, we have that

1. Standard deviation: $\rho(S(\lambda)) = \sqrt{\lambda^T \Sigma \lambda}$;
2. Value at Risk: $\rho(S(\lambda)) = \Phi^{-1}(\alpha)\sqrt{\lambda^T \Sigma \lambda}$;
3. Expected Shortfall: $\rho(S(\lambda)) = \dfrac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sqrt{\lambda^T \Sigma \lambda}$,

where $\phi(\,\cdot\,)$ and $\Phi^{-1}(\,\cdot\,)$ denote, respectively, the density and the inverse c.d.f. of univariate normal random variable. The importance of these results is the fact that in the multivariate Gaussian model all the three risk measures presented are proportional to the portfolio volatility and also simple to be calculated in closed form. This result was first discussed in the context of insurance risk management in [Panjer, 2001].

2.2.2 Euler allocation in a hierarchical structure

In this section we briefly extend the concept of Euler allocations to a bank structure divided, for example, in Business Units and Event types as suggested in the Basel II agreement for OpRisk (see Table 1.4).

Let us assume a bank has a structure given as in Figure 2.3, comprising of K Business Units (B.U.'s) and d_l Event Types (E.T's) in each of its B.U.'s ($l = 1, \ldots, K$). In this context we define $d = \sum_{l=1}^{K} d_l$ as the total number of cells for which capital should be allocated, $S = \sum_{i=1}^{d} X_i$ the bank loss and $X_{[l]} = \sum_{i_l=1}^{m} X_{i_l}$ for $l = 1, \ldots, K$ the loss in the l-th B.U..

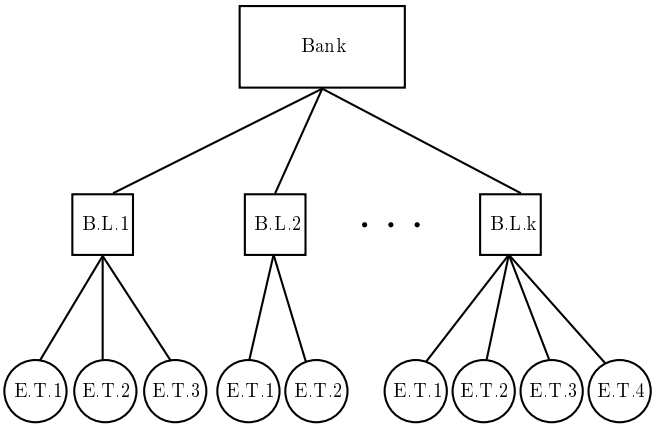

Figure 2.3: Hierarchical bank structure, with k B.U.'s.

Assuming the bank capital is given by $\mathrm{ES}_\alpha(S)$ (as in Definition 2.2.1) then the Euler principle states that the capital at the B.U. level should be given by $\mathbb{E}[X_{[l]} \,|\, S \geq \mathrm{VaR}_\alpha(S)]$ for

each B.U. $l = 1, \ldots, K$. To allocate the capital calculated for the l-th unit to its E.T.'s we can assume this capital is a (homogeneous) risk measure defined as

$$\rho(X_{[l]}) = \mathbb{E}[X_{[l]} \,|\, S \geq \mathrm{VaR}_\alpha(S)].$$

Then, following the Euler principle it is easy to check the allocation for the m-th E.T. in the l-th B.U. is given by

$$\mathbb{E}[X_m \,|\, S \geq \mathrm{VaR}_\alpha(S)].$$

The reader should note that the same allocations could have been derived "heuristically" in the following way. First the total capital $\mathrm{ES}_\alpha(S)$ is allocated directly to each of the d E.T.'s. Then for the l-th B.U. the capital is computed as the sum of the allocations at the d_l E.T.'s in this unit. Although the result would be the same, we emphasize the first method, as it is a direct application of the Euler principle (twice).

2.2.3 The relationship between Euler allocation and RORAC

To complete this section on capital allocation we also mention the relationship between the Euler allocation principle and other performance measures. For instance in [Tasche, 1999] and [Tasche, 2008], the Euler allocation principle is shown to be the only allocation principle that will also be Return On Risk-Adjusted Capital (RORAC) compatible, in a sense to be defined shortly.

If $\mu_i = E[X_i]$ denotes the expected return of the i-th component of a portfolio then the portfolio's Return On Risk-Adjusted Capital (RORAC) is defined as

$$RORAC(S) = \frac{E[S]}{\rho(S)} = \frac{\sum_{i=1}^{d} \mu_i}{\rho(S)}. \tag{2.11}$$

Similarly, the portfolio-related RORAC of the i-th asset in the portfolio is defined as

$$RORAC(X_i \,|\, S) = \frac{\mu_i}{\mathcal{A}_i^\rho}, \tag{2.12}$$

where $\mathcal{A}_i^\rho$ denotes the capital allocated (via a generic allocation principle) to the i-th asset based on a risk measure ρ.

Definition 2.2.8 (RORAC compatibility)**.** *Risk contributions are said to be* RORAC compatible *if there exist* $\epsilon_i > 0$, $i = 1, \ldots, d$ *such that* $\forall h \in (0, \epsilon_i)$

$$RORAC(X_i \,|\, S) > RORAC(S) \Longrightarrow RORAC(S + hX_i) > RORAC(S).$$

The importance of RORAC compatibility is that it provides us a way to create better portfolios. If we have that the portfolio-related RORAC of the i-th instrument (2.12) is greater than the total portfolio RORAC (2.11), then the portfolio RORAC can be improved by increasing the i-th position by a factor h.

Using the notion of RORAC compatibility, [Tasche, 1999] (Theorem 4.4) proved that for a given risk measure ρ the allocation principle will lead to RORAC compatibility if, and only if, it is the Euler allocation principle.

2.2.4 A critical view on the Euler allocation procedure in Insurance/OpRisk

The presentation of the Euler allocation principle given in Section 2.2 is based on the assumption that the portfolio for which the allocation process is being performed consists of a linear combination of random variables, see (2.6). When taking derivatives of the risk measure with respect to the weights we implicitly assume one can change the weights in the portfolio. Although these assumptions may not be appropriate in an insurance / OpRisk context (see [Mildenhall, 2004] and [Mildenhall, 2006] for an insurance related discussion) [Boonen et al., 2016] recently provide arguments to suport the use of the Euler principle, at least as an approximation to the "correct" problem.

As mentioned above, the main criticism to the Euler principle in an insurance context (and the same is also valid for OpRisk) is the "linearity assumption". A good example is provided in [Mildenhall, 2006], when studying two different "Motor Hull" portfolios: the first one contains a single policy and the other is very large. In his words, "*for the single policy the probability of no claims in one year is around* 90%. *For the large portfolio the probability of no claims will be very close to zero. Thus there cannot be a random variable* X *such that the single policy losses have distribution* x_1X *and the [large] portfolio has distribution* x_2X".

To overcome the "linearity problem", in contrast to (2.6), [Boonen et al., 2016] introduce the following total loss process, defined in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}_{0\leq t\leq T})$ with $\mathcal{F} = \mathcal{F}_T$,

$$S(\boldsymbol{\lambda}) = \sum_{i=1}^{d} X_i(\lambda_i), \quad \text{for} \quad \lambda_i \in \Lambda = [0, T]$$

with a current (base-line) exposure of $\boldsymbol{\lambda} = \mathbf{1}_d = (1, \ldots, 1)$. Losses in the i-th LoB are then defined as

$$X_i(\lambda_i) = Y_i(\lambda_i) + \lambda_i Z_i, \tag{2.13}$$

where $\{Y_i(\lambda)\}_{0\leq\lambda\leq T}$ are $\mathcal{F}_\lambda$-adapted independent increasing Lévy processes and each of the Z_i's is $\mathcal{F}_T$ measurable, independent of $Y_i(\lambda)$. Under the representation in (2.13) the first term should be viewed as the (non-linear) insurance risk component and the second as common shocks that affect all the claims in the i-th LoB.

Using the nomenclature from [Boonen et al., 2016] we call any function $g : \Lambda \to \mathbb{R}$ a *fuzzy game*. Note that the risk measure function $r_\rho(\boldsymbol{\lambda}) : \Lambda \to \mathbb{R}$ is a fuzzy game, but the latter may not be defined through a risk measure. Moreover, a vector $\mathbf{d} \in \mathbb{R}^d$ is called an *allocation of the fuzzy game* if $\sum_{i=1}^{d} d_i = g(\mathbf{1}_d)$. Another important concept is the *core of a fuzzy game*, which is defined as

$$\mathcal{C}(g) = \left\{ \mathbf{d} \in \mathbb{R}^d \ : \ \mathbf{d} \text{ is an allocation of } g \text{ and } \sum_{i=1}^{d} \lambda_i d_i \leq g(\boldsymbol{\lambda}), \ \forall \boldsymbol{\lambda} \in [0,1]^d \right\}.$$

The core of the game should be seen as the set of allocations that do not give an incentive to split the portfolio, as the risk contribution of any sub-portfolio is less than its stand-alone

capital. In [Aubin, 2007] it has been proved that if g is subadditive then the Euler allocation is the unique element of the core, an argument explored by [Denault, 2001] to justify its use.

Given a risk measure ρ and the structure in (2.13) the overall capital to be held, as a function of the exposures $\boldsymbol{\lambda}$ is given by the following fuzzy game:

$$r_\rho(\boldsymbol{\lambda}) = \rho\left(\sum_{i=1}^{d} X_i(\lambda_i)\right), \quad \text{for} \quad \lambda_i \in \Lambda = [0, T],$$

and in this case r_ρ is not necessarily positive homogeneous and, therefore, the Euler principle is not directly applicable.

To circumvent this problem, [Boonen et al., 2016] introduce a "linearized version" of the fuzzy game r_ρ, defined as

$$\widetilde{r}_\rho(\boldsymbol{\lambda}) = \rho\left(\sum_{i=1}^{d} \lambda_i X_i(1)\right)$$

and then prove (see Corollary 3.8 of the original paper) that the Euler allocation for the fuzzy game $\widetilde{r}_\rho$ belongs to the core of r_ρ. Note that, as previously discussed, $\widetilde{r}_\rho$ is the unique element in the core of $\widetilde{r}_\rho$, but in [Boonen et al., 2016, Section 3.3] it is shown that $\mathcal{C}(r_\rho)$ is not necessarily single-valued.

Therefore, paraphrasing [Boonen et al., 2016], "allocations used in practice [*i.e., using the Euler principle and linearity assumptions*] are based on the 'wrong game' $\widetilde{r}_\rho$, rather than the 'game actually played', r_ρ. But our results show that this is actually not a serious problem, since via the Euler allocation of $\widetilde{r}_\rho$, we end up with an element of the core of r_ρ".

Chapter 3

Monte Carlo methods

In the previous chapter we saw that, under the Euler principle, for both the value at risk (VaR) and the expected shortfall (ES) marginal risk allocations are given by conditional expectations. For a generic risk model, i.e., a combination of marginal distributions and a copula, these conditional expectations cannot be calculated in closed form, so numerical approximations have to be used.

Some exceptions do exist, though. Models where allocations can be explicitly calculated include, for example, the multivariate Gaussian model from Section 2.2.1 (extended to the case of multivariate elliptical distributions in [Landsman and Valdez, 2003] and [Dhaene et al., 2008]), the multivariate gamma model of [Furman and Landsman, 2005], the combination of the Farlie-Gumbel-Morgenstern (FGM) copula and (mixtures of) exponential marginals from [Bargès et al., 2009] or (mixtures of) Erlang marginals [Cossette et al., 2013], the multivariate Pareto-II from [Asimit et al., 2013].

In this chapter we present some background on Monte Carlo (MC) methods and set up the framework for the novel simulation based algorithms for capital allocation developed in Chapters 4 and 6. In particular, we review the recent class of *pseudo-marginal* Markov Chain Monte Carlo (MCMC) and Sequential Monte Carlo (SMC) algorithms, the latter being the basis of the algorithms proposed.

Throughout the chapter we assume the marginal loss processes $X_1, \dots, X_d$ are continuous random variables, implying that all marginal inverse c.d.f.'s (quantile functions) F_i^{-1} are well defined and also continuous. Moreover, from Sklar's theorem we have that the joint c.d.f. and p.d.f. of the vector $\mathbf{X}$ are, respectively,

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d] = C(F_1(x_1), \dots, F_d(x_d)) \text{ and}$$

$$f_{\mathbf{X}}(\mathbf{x}) = c\big(F_1(x_1), \dots, F_d(x_d)\big) \prod_{i=1}^{d} f_i(x_i),$$

where C and c are the copula and copula density, respectively.

From Proposition 2.2.6 we see both the allocation based on VaR and on ES take the form of a conditional expectation, such as

$$\mathcal{A}_i^{\rho}(X_i) = \mathbb{E}[h(\mathbf{X}) \,|\, g(\mathbf{X}) \in A], \tag{3.1}$$

for the following choices of h, g and A:

1. For $\rho = \text{VaR}$: $h(\mathbf{X}) = X_i,\ g(\mathbf{X}) = \sum_{i=1}^d X_i,\ A = \Big[\text{VaR}_\alpha(\sum_{i=1}^d X_i),\ \text{VaR}_\alpha(\sum_{i=1}^d X_i)\Big]$;
2. For $\rho = \text{ES}$: $h(\mathbf{X}) = X_i,\ g(\mathbf{X}) = \sum_{i=1}^d X_i,\ A = \Big[\text{VaR}_\alpha(\sum_{i=1}^d X_i),\ +\infty\Big)$.

Therefore, if we are able to generate values from

$$\pi(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x} \,|\, g(\mathbf{X}) \in A) \tag{3.2}$$

then these samples can be used to approximate (3.1). In the MC literature the density defined in (3.2) is usually referred to as the *target density*, while the samples are sometimes called *particles*.

3.1 Simple Monte Carlo and rejection sampling

The idea behind a *simple* (or naive or crude) Monte Carlo simulation scheme is to generate a set of independent and identically distributed (i.i.d.) samples from the target distribution, which can then be used to approximate the required expectations.

When a set of i.i.d. samples $\{\mathbf{x}^{(j)}\}_{j=1}^N$ from the target distribution π is available, the MC estimator of $\mathbb{E}_\pi[\varphi(\mathbf{X})]$, the expectation under π of a generic test function φ, is given by

$$\widehat{\varphi}_{MC}^N = \frac{1}{N}\sum_{j=1}^N \varphi(\mathbf{x}^{(j)}).$$

Provided that $\mathbb{E}_\pi[\varphi(\mathbf{X})]$ is finite, the strong law of large numbers (SLLN) for i.i.d. variables (see, e.g. [Shiryaev, 1996, Chapter III]) ensures that

$$\lim_{N\to+\infty} \widehat{\varphi}_{MC}^N \xrightarrow{a.s.} \mathbb{E}_\pi[\varphi(\mathbf{X})]. \tag{3.3}$$

In other words, as long as the test function have finite expectation under π the MC method is guaranteed to converge almost surely to the correct quantity.

Moreover, if the variance of the test function under the target distribution is finite then a central limit theorem (CLT) holds (see [Shiryaev, 1996, Chapter III]), ensuring the distribution of the estimator approaches the true quantity when the sample size goes to infinity:

$$\lim_{N\to+\infty} \sqrt{N}(\widehat{\varphi}_{MC}^N - \mathbb{E}_\pi[\varphi(\mathbf{X})]) \xrightarrow{d} N(0, \text{Var}_\pi(\varphi(\mathbf{X}))),$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Remark 3.1.1 (Inversion method)**.** *Whenever the target density π is defined over the real numbers and its c.d.f. F is invertible in closed form, one can use the fact that*

$$U \sim Unif[0,1] \Rightarrow X = F^{-1}(U) \sim \pi$$

to generate i.i.d. samples from π.

3.1.1 Rejection sampling

If samples cannot be generated directly from the target distribution one possibility is to use *rejection sampling*. In short, this simple method makes use of an auxiliary distribution which can be easily sampled from and then accepts (or reject, hence the name) the samples with a specific probability.

As before, let us assume our interest is to generate samples from the target distribution π but now we assume we have access to another density, μ such that $\pi(\mathbf{x}) \leq M\mu(\mathbf{x})$ for $1 < M < +\infty$, for all $\mathbf{x}$. In simple words, in order to draw samples from π we can sample $\mathbf{x}$ from μ and accept it as a sample from the target distribution with probability $\frac{\pi(\mathbf{x})}{M\mu(\mathbf{x})}$. This procedure is formalized in Algorithm 1.

Lemma 3.1.2. *The rejection sampling algorithm (Algorithm 1) produces a variable Y distributed as π.*

Proof. The proof presented below is based on [Robert and Casella, 2004, Lemma 2.3.1].

The cdf of $\mathbf{Y}$ is given by

$$\mathbb{P}[\mathbf{Y} \leq \boldsymbol{y}] = \mathbb{P}\left[\mathbf{X} \leq \boldsymbol{y} \,|\, U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})}\right] = \frac{\mathbb{P}\left[\mathbf{X} \leq \boldsymbol{y},\, U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})}\right]}{\mathbb{P}\left[U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})}\right]}.$$

To show the above cdf is the same as the cdf of$\mathbf{X}$, note the denominator is the unconditional acceptance probability of the rejection method, which is the proportion of proposed samples that are accepted:

$$\begin{aligned}
\mathbb{P}\left[U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})}\right] &= \mathbb{E}\left[\mathbb{P}\left[U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})} \,\middle|\, \mathbf{X} = \mathbf{x}\right]\right] \\
&= \mathbb{E}_{\mu}\left[\frac{\pi(\mathbf{x})}{M\mu(\mathbf{x})}\right] \\
&= \int \frac{\pi(\mathbf{x})}{M\mu(\mathbf{x})}\mu(\mathbf{x})d\mathbf{x} \\
&= \frac{1}{M}\int \pi(\mathbf{x})d\mathbf{x} \\
&= \frac{1}{M}.
\end{aligned}$$

Similarly, for the numerator we have that

$$\mathbb{P}\left[\mathbf{X} \leq \boldsymbol{y},\, U \leq \frac{\pi(\mathbf{X})}{M\mu(\mathbf{X})}\right] = \frac{1}{M}\mathbb{P}[\mathbf{X} \leq \boldsymbol{y}].$$

□

Remark 3.1.3. *When the target distribution is the truncated version of another distribution, i.e., $\pi(\boldsymbol{x}) = \mu(\boldsymbol{x}\,|\, g(\boldsymbol{x}) \in A) = \frac{\mu(\boldsymbol{x})\mathbb{1}_{\{g(\boldsymbol{x})\in A\}}}{\mathbb{P}[g(\boldsymbol{X}) \in A]}$ we have that $M = \frac{1}{\mathbb{P}[g(\boldsymbol{X})\in A]}$. From the above results, the expected number of iterations needed in order to generate one sample from π is inversely proportional to the probability of the event, which can be prohibitive if the probability is "too small".*

Algorithm 1: Rejection sampling algorithm.

Inputs: Auxiliary density μ such that $\mu(\mathbf{x}) \geq \frac{\pi(\mathbf{x})}{M}$;

Sample $\mathbf{X} \sim \mu$;

Sample $U \sim U[0,1]$;

if $U \leq \dfrac{\pi(\boldsymbol{x})}{M\mu(\boldsymbol{x})}$ **then**

| Accept $\mathbf{Y} = \mathbf{X}$ as a sample from π.

else

| Reject the sample and go to step 1.

end

Result: $\mathbf{Y}$ a sample from π.

3.2 Importance Sampling

If a "good" auxiliary distribution cannot be designed, for example, the target distribution is defined in a high dimensional space or in a constrained region, the rejection sampling scheme of the previous section can be extremely inefficient, as many samples would be rejected. Importance sampling (IS) tries to overcome this problem using all the samples simulated from the auxiliary distribution (also known as the importance distribution) each of which with an importance weight.

For a generic measurable function $\varphi(\mathbf{x})$ the cornerstone identity for IS is

$$\mathbb{E}_\pi[\varphi(\mathbf{X})] = \int \varphi(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} = \int \varphi(\mathbf{x})\frac{\pi(\mathbf{x})}{\mu(\mathbf{x})}\mu(\mathbf{x})d\mathbf{x} = \int \varphi(\mathbf{x})w(\mathbf{x})\mu(\mathbf{x})d\mathbf{x} = \mathbb{E}_\mu[\varphi(\mathbf{x})w(\mathbf{x})], \tag{3.4}$$

which is valid for any importance density μ such that its support is larger than π's, i.e., $supp(\mu) \supset supp(\pi)$.

Given a set of samples $\{\mathbf{x}^{(j)}\}_{j=1}^N$ from μ we construct the IS estimate as

$$\widehat{\varphi}_{IS}^N = \frac{1}{N}\sum_{j=1}^N w(\mathbf{x}^{(j)})\varphi(\mathbf{x}^{(j)})$$

which, analogously to (3.3) converges to $\mathbb{E}_\mu[\varphi(\mathbf{x})w(\mathbf{x})]$ when $N \longrightarrow +\infty$. Therefore, from (3.4),

$$\lim_{N\longrightarrow+\infty} \widehat{\varphi}_{IS}^N = \mathbb{E}_\pi[\varphi(\mathbf{X})]$$

and a similar CLT result is also available.

In many cases, including the ones discussed in this work, the target distribution π cannot be evaluated due to an intractable normalizing constant. In this context we have

$$\pi(\mathbf{x}) = \mathcal{Z}^{-1}\gamma(\mathbf{x}) \propto \gamma(\mathbf{x}),$$

where $\mathcal{Z}$ is unknown. Therefore, the IS weights can only be computed in its *self-normalized* form, defined as

$$W(\mathbf{x}^{(j)}) = \frac{w(\mathbf{x}^{(j)})}{\sum_{k=1}^N w(\mathbf{x}^{(k)})},$$

for $w(\mathbf{x}) = \dfrac{\pi(\mathbf{x})}{\mu(\mathbf{x})}$, as in (3.4).

As shown in [Geweke, 1989], the estimator

$$\widehat{\varphi}^N_{IS,sn} = \sum_{j=1}^{N} W(\mathbf{x}^{(j)})\varphi(\mathbf{x}^{(j)}),$$

based on the self-normalized IS weights is also consistent and satisfies a CLT result, but as the next result shows there is a price to be paid: the self-normalized IS estimator is biased for finite samples, but the bias, in practice, in usually very small. Also, the estimator is asymptotically unbiased. For a proof the reader is referred to, for example, [Johansen and Evers, 2007, Proposition 3.2.1].

Proposition 3.2.1. *The bias and variance of the two IS estimators are given by*

(a) $\mathbb{E}_\mu[\widehat{\varphi}^N_{IS}(\boldsymbol{X})] = \mathbb{E}_\pi[\varphi(\boldsymbol{X})]$

(b) $Var_\mu(\widehat{\varphi}^N_{IS}(\boldsymbol{X})) = \dfrac{Var_\mu(w(\boldsymbol{X})\varphi(\boldsymbol{X}))}{N}$

(c) $\mathbb{E}_\mu[\widehat{\varphi}^N_{IS,sn}(\boldsymbol{X})] = \mathbb{E}_\pi[\varphi(\boldsymbol{X})] + \dfrac{\mathbb{E}_\pi[\varphi(\boldsymbol{X})]\, Var_\mu(w(\boldsymbol{X})) - Cov_\mu(w(\boldsymbol{X}),\, w(\boldsymbol{X})\varphi(\boldsymbol{X}))}{N} + O(N^{-2})$

(d) $Var_\mu(\widehat{\varphi}^N_{IS,sn}(\boldsymbol{X})) = \dfrac{Var_\mu(w(\boldsymbol{X})\varphi(\boldsymbol{X})) - 2\mathbb{E}_\pi[\varphi(\boldsymbol{X})]Cov_\mu(w(\boldsymbol{X}),\, w(\boldsymbol{X})\varphi(\boldsymbol{X})) + \mathbb{E}_\pi[\varphi(\boldsymbol{X})]^2\, Var_\mu(w(\boldsymbol{X}))}{N} + O(N^{-2})$

Remark 3.2.2. *Although the self-normalized estimator $\widehat{\varphi}^N_{IS,sn}$ is asymptotically biased, its variance can be lower than the unbiased estimator, $\widehat{\varphi}^N_{IS}$.*

3.3 Markov Chain Monte Carlo (MCMC) methods

The aim of this section is to provide a brief overview of Markov Chain Monte Carlo algorithms, which are an alternative method to generate samples from a complex distribution. There is a vast literature on MCMC methods in Statistics and we refer the reader to [Gamerman and Lopes, 2006] and [Robert and Casella, 2004] for an in-depth introduction to the topic.

3.3.1 Metropolis Hastings

In order to generate samples from a target distribution denoted by $\pi(\mathbf{x}) \propto \gamma(\mathbf{x})$, the Metropolis-Hastings (MH) algorithm (first proposed in [Metropolis et al., 1953] and generalized in [Hastings, 1970]) simulates a Markov chain with stationary distribution equals to π. Given the unnormalized target distribution γ, an initial state $\mathbf{x}_0$ and a proposal function $q(\mathbf{x}' \,|\, \mathbf{x})$, after a sufficiently large number of iterations, MH generates a sequence of states $\mathbf{x}_1, \ldots, \mathbf{x}_T$ from a sequence of distributions that converges to π. The pseudo-code is presented in Algorithm 2.

Note that, differently from IS, the MH has a rejection step, implying that not all simulated values will be part of the Markov chain. Another difference is the fact that the samples generated

through MH are dependent (the method, indeed, creates a Markov Chain).

Algorithm 2: Generic Metropolis Hastings (MH) algorithm.

Inputs: Unnormalized target distribution $\gamma(\mathbf{x})$, initial state $\mathbf{x}_0$, proposal density $q(\mathbf{x}' \,|\, \mathbf{x})$ and total number of iterations T;

for $t = 1, \ldots, T$ **do**

Propose $\mathbf{x}' \sim q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})$;

Calculate $a = \frac{\gamma(\mathbf{x}')q(\mathbf{x}_{t-1} \,|\, \mathbf{x}')}{\gamma(\mathbf{x}_{t-1})q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})}$;

Set $\mathbf{x}_t = \mathbf{x}'$ with probability $\min\{1, a\}$, i.e., ;

1. Draw $r \sim Unif[0, 1]$;
2. Set $\mathbf{x}_t = \mathbf{x}'$ if $r < a$.

end

Result: Samples $\{\mathbf{x}_t\}_{t=0}^T$ from a Markov chain with invariant distribution $\pi(\mathbf{x}) \propto \gamma(\mathbf{x})$.

3.3.2 Gibbs sampler

The Gibbs sampling algorithm [Geman and Geman, 1984] resamples each coordinate of the vector $\mathbf{x}$ from their full conditional distributions $\pi(x_i \,|\, \mathbf{x}_{-i})$, where $\mathbf{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$. Alternatively, we can write the Gibbs sampling updates as the following MH proposal

$$q(\mathbf{x}' \,|\, \mathbf{x}) = \gamma(x_i' \,|\, \mathbf{x}_{-i}) \mathbb{1}_{\{\mathbf{x}'_{-i} = \mathbf{x}_{-i}\}},$$

where $\mathbb{1}_{\{\mathbf{x}'_{-i} = \mathbf{x}_{-i}\}}$ denotes an indicator function ensuring all components of $\mathbf{x}$ are fixed, apart from x_i. As the acceptance probability for this proposal is always one, the Gibbs sampler does not suffer from the inefficiency of rejecting simulated values. The algorithm is summarized in Algorithm 3, where we use the notation $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,d})$.

Algorithm 3: Generic Gibbs sampler algorithm.

Inputs: Full conditional p.d.f.'s $\pi(x_i \,|\, \mathbf{x}_{-i})$, for $i = 1, \ldots, d$, initial state $\mathbf{x}_0$ and total number of iterations T ;

for $t = 1, \ldots, T$ **do**

for $i = 1, \ldots, d$ **do**

Sample $x_{t,i} \sim \pi(x_{t,i} \,|\, \mathbf{x}_{t-1,-i})$;

end

end

Result: Samples $\{\mathbf{x}_t\}_{t=0}^T$ from a Markov chain with invariant distribution $\pi(\mathbf{x}) \propto \gamma(\mathbf{x})$.

One of the main benefits of the Gibbs sampling algorithm is the fact that there are not free parameters, so no tuning is needed and the algorithm can be applied in an automatic fashion – as long as it is possible to sample from the full conditionals.

Remark 3.3.1. *As with MH, Gibbs samplers generate a Markov chain whose samples approximate the target distribution only in the limit and care should be taken in order to assess the*

convergence of these chains in practice. For a discussion on convergence diagnostics of MCMC methods the reader is referred to [Gamerman and Lopes, 2006, Section 5.4].

3.3.3 Slice sampling

The last algorithm discussed in this section, the slice sampler of [Neal, 2003] is, as the rejection sampling algorithm, based on the simulation of an auxiliary variable. Similarly to the Gibbs sampler it has the advantage of no rejections.

For a univariate distribution $\pi(x) \propto \gamma(x)$ the slice sampling algorithm is described in Algorithm 4.

Algorithm 4: Generic slice sampling algorithm.

Inputs: Unnormalized target density $\gamma(x)$ and initial state x_0;

Sample $y \sim Unif[0, \gamma(x_0)]$;

Define the interval $I = \{x \ : \ \gamma(x) > y\}$;

Sample $x \sim Unif(I)$;

Result: One sample x from $\pi(x) \propto \gamma(x)$.

Remark 3.3.2. *Another prominent member of the class of Monte Carlo algorithms based in auxiliary variables is the HMC (where the acronym stands for either Hamiltonian Monte Carlo or Hybrid Monte Carlo, depending on the authors!), which is not discussed here. The interested reader is referred to [Duane et al., 1987] and [Neal, 2011].*

3.3.4 Pseudo-marginal MCMC

Although the MH algorithm is an extremely versatile method to sample from a target density π, if this density is not known in closed form (or if it is computationally too expensive to calculate) then MH in its original form cannot be directly applied. One example of such scenario arises in Section 6.2.2, when the target density is defined conditional to a nuisance parameter. The problem of intractable target distributions also appears when studying state-space models, where the likelihood is defined through a high-dimensional integral (see, e.g., [Andrieu et al., 2010]).

Over the past decades promising research has been developed towards algorithms that can circumvent the calculation of the exact MH acceptance probability. The idea of using unbiased estimators within MH algorithms was first suggested in the Physics literature by [Kennedy and Kuti, 1985] and further explored in [Lin et al., 2000]. These ideas were then introduced to the Bayesian statistics community in [Beaumont, 2003] and [Andrieu and Roberts, 2009] where it was given the *pseudo-marginal* name. Since then many theoretical results have been obtained for these algorithms, including qualitative ones in [Andrieu and Roberts, 2009] and [Andrieu and Vihola, 2015] and quantitative ones in [Pitt et al., 2012], [Sherlock et al., 2015], [Doucet et al., 2015] and [Deligiannidis et al., 2015].

The idea of replacing an unknown density by a positive and unbiased estimate is in the core of many recently proposed algorithms, such as the Particle Markov Chain

Monte Carlo (PMCMC) of [Andrieu et al., 2010], the Sequential Monte Carlo Squared (SMC^2) of [Chopin et al., 2013] and [Fulop and Li, 2013] (see also the island particle filter of [Vergé et al., 2015]) and the Importance Sampling Squared (IS^2) of [Tran et al., 2014]. In the context of Sequential Monte Carlo algorithms this argument first appeared as a brief note in Rousset and Doucet's comments of [Beskos et al., 2006], where it reads that "(...) a straightforward argument shows that it is not necessary to know $w_k(X^{(i)}_{t_0:t_k})$ [*the weights*] exactly. Only an unbiased positive estimate $\hat{w}_k(X^{(i)}_{t_0:t_k})$ of $w_k(X^{(i)}_{t_0:t_k})$ is necessary to obtain asymptotically consistent SMC estimates under weak assumptions".

Let us denote by $\widehat{\gamma}(\mathbf{x})$ an estimate of the unnormalized target density γ. As we are assuming γ cannot be directly evaluated, one strategy to use the MH is simply to substitute γ by $\widehat{\gamma}$ in the acceptance ratio, i.e., to calculate something like

$$\widehat{a} = \frac{\widehat{\gamma}(\mathbf{x}')q(\mathbf{x}_{t-1} \,|\, \mathbf{x}')}{\widehat{\gamma}(\mathbf{x}_{t-1})q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})}.$$

To prove that this is actually a valid MH we need to provide some details on how the estimator $\widehat{\gamma}$ is calculated and some of its necessary properties. Let us assume the estimator $\widehat{\gamma}$ is calculated using a random vector $\mathbf{u} \sim p(\mathbf{u})$, i.e., we approximate $\gamma(\mathbf{x})$ by $\widehat{\gamma}(\mathbf{x} \,|\, \mathbf{u})$. Furthermore, we also assume that the estimator is non-negative, $\widehat{\gamma}(\mathbf{x} \,|\, \mathbf{u}) \geq 0,\ \forall(\mathbf{x},\, \mathbf{u})$, and unbiased, in the following sense

$$\gamma(\mathbf{x}) = \mathbb{E}_{\mathbf{U}}[\widehat{\gamma}(\mathbf{x} \,|\, \mathbf{U})]. \tag{3.5}$$

As the slice sampler, the pseudo-marginal MH operates in an extended space, in this case the space of $(\mathbf{x},\, \mathbf{u})$, where the target density is defined as

$$\gamma(\mathbf{x},\, \mathbf{u}) = \widehat{\gamma}(\mathbf{x} \,|\, \mathbf{u})p(\mathbf{u}).$$

Note that, due to the unbiasedness assumption (3.5), $\gamma(\mathbf{x},\, \mathbf{u})$ has $\gamma(\mathbf{x})$ as marginal. Taking the proposal density in the extended space as

$$\begin{aligned} q(\mathbf{x}',\, \mathbf{u}' \,|\, \mathbf{x}_{t-1},\, \mathbf{u}_{t-1}) &= q(\mathbf{x}' \,|\, \mathbf{u}_{t-1},\, \mathbf{x}_{t-1})q(\mathbf{u}' \,|\, \mathbf{x}_{t-1},\, \mathbf{u}_{t-1}) \\ &= q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})p(\mathbf{u}') \end{aligned}$$

the MH acceptance ratio becomes

$$\begin{aligned} a &= \frac{\gamma(\mathbf{x}',\, \mathbf{u}')q(\mathbf{x}_{t-1},\, \mathbf{u}_{t-1} \,|\, \mathbf{x}',\, \mathbf{u}')}{\gamma(\mathbf{x}_{t-1},\, \mathbf{u}_{t-1})q(\mathbf{x}',\, \mathbf{u}' \,|\, \mathbf{x}_{t-1},\, \mathbf{u}_{t-1})} \\ &= \frac{\widehat{\gamma}(\mathbf{x}' \,|\, \mathbf{u}')q(\mathbf{x}_{t-1} \,|\, \mathbf{x}')}{\widehat{\gamma}(\mathbf{x}_{t-1} \,|\, \mathbf{u}_{t-1})q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})}, \end{aligned}$$

which does not depend on the unknown density γ. Therefore, the above acceptance ratio can be used in an MH algorithm (marginally) targeting γ. This algorithm, named the pseudo-marginal MH, is outlined in Algorithm 5.

3.3.5 Some comments

As MCMC algorithms can be used to construct Markov chains with a pre-specified target distribution it can, in theory, be used to generate samples from the highly constrained target

Algorithm 5: Generic pseudo-marginal Metropolis Hastings (MH) algorithm.

Inputs: Positive and unbiased estimate for the unnormalized target $\widehat{\gamma}(\mathbf{x} \,|\, \mathbf{u})$, initial state $\mathbf{x}_0$, proposal density $q(\mathbf{x}' \,|\, \mathbf{x})$ and total number of iterations T;

for $t = 1, \ldots, T$ **do**

Sample $\mathbf{u}' \sim p(\mathbf{u}')$;

Propose $\mathbf{x}' \sim q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})$;

Calculate $a = \frac{\widehat{\gamma}(\mathbf{x}' \,|\, \mathbf{u}')q(\mathbf{x}_{t-1} \,|\, \mathbf{x}')}{\widehat{\gamma}(\mathbf{x}_{t-1} \,|\, \mathbf{u}_{t-1})q(\mathbf{x}' \,|\, \mathbf{x}_{t-1})}$;

Set $(\mathbf{x}_t, \mathbf{u}_t) = (\mathbf{x}', \mathbf{u}')$ with probability $\min\{1, a\}$, i.e., ;

1. Draw $r \sim Unif[0, 1]$;
2. Set $(\mathbf{x}_t, \mathbf{u}_t) = (\mathbf{x}', \mathbf{u}')$ if $r < a$.

end

Result: Samples $\{\mathbf{x}_t\}_{t=0}^{T}$ from a Markov chain with invariant distribution $\pi(\mathbf{x}) \propto \gamma(\mathbf{x})$.

distribution defined in (3.2). Each one of the algorithms discussed above present different challenges.

As presented in Section 3.3.3 the slice sampling algorithm is only applicable to unidimensional distributions, which may be known up to a proportionality constant. For the target distributions defined in (3.2) the marginals (which are, indeed, the distributions to be used in the allocation problem) take the following form:

$$\begin{aligned}
\pi(x_i) &= \int \pi(\mathbf{x}) d\mathbf{x}_{-i} \\
&= \int f(\mathbf{x} \,|\, g(\mathbf{x}) \in A) d\mathbf{x}_{-i} \\
&= \frac{1}{1-\alpha} \int f(x_i) f(\mathbf{x}_{-i} \,|\, x_i) \mathbb{1}_{\{g(\mathbf{x}) \in A\}} d\mathbf{x}_{-i},
\end{aligned}$$

which cannot, in general, be calculated in close form. A similar problem arises when one tries a simple implementation of the Gibbs sampling algorithm of Section 3.3.2, since the full conditionals $\pi(x_i \,|\, \mathbf{x}_{-i}) = \frac{\pi(\mathbf{x})}{\pi(\mathbf{x}_{-i})}$ cannot be calculated either.

Due to its wide applicability the MH algorithm is, in theory, suitable for sampling from the target distributions in (3.2) but, in practice, in order to avoid too many rejections, the proposal distribution should be such that most of its mass is concentrated in the rare set $\{g(\mathbf{x}) \in A\}$, which is a non trivial task. Note that this is the same issue that arises when one tries to, naively, use IS to sample from π, as most of the samples could end up having zero weight.

Next we present a class of algorithms that extends IS solutions to sequential settings, known in the statistics literature as Sequential Monte Carlo (SMC) methods, making a clear distinction between SMC and SMC sampler algorithms. Although based on IS techniques, SMC methods also make use of MCMC algorithms but differently from the above discussion the MCMC kernels in SMC are applied to samples that are already in the stationary distribution (i.e., the constrained region), as discussed in Section 3.4.1.

3.4 Sequential Monte Carlo methods (SMC)

In this section we introduce the general class of algorithms known as Sequential Monte Carlo (SMC) and an important variant for rare-event simulation, the SMC Samplers classes of methods. This family of Monte Carlo algorithms has been developed to approximate sequences of integrals constructed from a sequence of probability density functions. Of course, adjustments are possible when the interest lies only in one distribution, such as the terminal distribution in a sequence of increasingly rare events, an idea explored in Chapter 4 (see also Section 3.6).

Historically SMC methods emerged out of the fields of engineering, probability and statistics and variants of the methods sometimes appear under the names of particle filtering or interacting particle systems e.g. [Ristic et al., 2004], [Doucet et al., 2001], [Del Moral, 2004]. Their theoretical properties have been extensively studied in [Crisan and Doucet, 2002], [Del Moral, 2004], [Chopin, 2004], [Künsch, 2005].

For a recent survey in the topic, with focus on economics, finance and insurance applications the reader is referred to [Creal, 2012] and [Del Moral et al., 2013]. One of the most successful applications of SMC methods in finance is in the field of option pricing: American options are discussed in [Rambharat and Brockwell, 2010], Asian options in [Jasra and Del Moral, 2011], Barrier options in [Shevchenko and Del Moral, 2015] and, more recently, [Sen et al., 2016] revisit the pricing of barrier options and also introduce algorithms for Target Accumulation Redemption Notes (TARNs). In the context of credit portfolio losses [Carmona et al., 2009] and [Carmona and Crépey, 2010] discuss the use of SMC algorithms to estimate small default probabilities.

The general context of a standard SMC method is that one wants to approximate a (often naturally occurring) sequence of probability density functions (p.d.f.'s) $\{\widetilde{\pi}_t\}_{t\geq 1}$ such that the support of every function in this sequence is defined as $supp(\widetilde{\pi}_t) = E_t$ and the dimension of E_t forms an increasing sequence, i.e., $dim(E_{t-1}) < dim(E_t)$. For example, the reader can think of $E_1 = \mathbb{R}^d, \ldots, E_t = \mathbb{R}^{d\times t}$, which will be precisely the sequence to be used throughout this work.

We may also assume that $\widetilde{\pi}_t$ is only known up to a normalizing constant,

$$\widetilde{\pi}_t(\mathbf{x}_{1:t}) = \mathcal{Z}_t^{-1}\widetilde{\gamma}_t(\mathbf{x}_{1:t}),$$

where $\mathbf{x}_{1:t} = (\mathbf{x}_1, \ldots, \mathbf{x}_t) \in E_t = \mathbb{R}^{d\times t}$. As in Section 3.5, the approximation for $\widetilde{\pi}_t$ is given by a weighted sum of random samples (also known as "particles").

Procedurally, we initialize the algorithm sampling a set of N particles from the distribution $\widetilde{\pi}_1$ and set the normalized weights to $W_1^{(j)} = 1/N$, for all $j = 1, \ldots, N$. If it is not possible to sample directly from $\widetilde{\pi}_1$, one should sample from an importance distribution $\widetilde{q}_1$ and calculate its weights accordingly (see Algorithm 6). Then the particles are sequentially propagated thorough each distribution $\widetilde{\pi}_t$ in the sequence, via three main processes: mutation, correction (incremental importance weighting) and resampling. In the first step (mutation) we propagate particles from time $t-1$ to time t and in the second one (correction) we calculate the new importance weights of the particles.

Algorithm 6: Standard SMC algorithm.

Inputs: IS density $\widetilde{q}_1$, (forward) mutation kernels $\left\{K_t(\mathbf{x}_{t-1}, \mathbf{x}_t)\right\}_{t=1}^{T}$;

for $j = 1, \ldots, N$ **do**

Sample $\mathbf{x}_1^{(j)}$ from $\mathbf{X}_1 \sim \widetilde{q}_1(\,\cdot\,)$ *(Mutation step)*;

Calculate the weights $w_1^{(j)} = \frac{\widetilde{\gamma}_1(\mathbf{x}_1^{(j)})}{\widetilde{q}_1(\mathbf{x}_1^{(j)})}$;

end

Calculate the normalized weights $W_1^{(j)} = \frac{w_1^{(j)}}{\sum_{j=1}^{N} w_1^{(j)}}$ *(Correction step)*;

for $t = 2, \ldots, T$ **do**

for $j = 1, \ldots, N$ **do**

Sample $\mathbf{x}_t^{(j)}$ from $\mathbf{X}_t \big| \mathbf{X}_{t-1} = \mathbf{x}_{t-1}^{(j)} \sim K_t(\mathbf{x}_{t-1}^{(j)}, \,\cdot\,)$ *(Mutation step)*;

Create the vector $\mathbf{x}_{1:t}^{(j)} = (\mathbf{x}_{1:(t-1)}^{(j)}, \mathbf{x}_t^{(j)})$;

Calculate the weights $w_t^{(j)} = \dfrac{\widetilde{\gamma}_t(\mathbf{x}_{1:t}^{(j)})}{\widetilde{q}_t(\mathbf{x}_{1:t}^{(j)})} = w_{t-1}^{(j)} \underbrace{\dfrac{\widetilde{\gamma}_t(\mathbf{x}_{1:t}^{(j)})}{\widetilde{\gamma}_{t-1}(\mathbf{x}_{1:t-1}^{(j)}) K_t(\mathbf{x}_{t-1}^{(j)}, \mathbf{x}_t^{(j)})}}_{\text{incremental weight: } \widetilde{\alpha}(x_{1:t}^{(j)})}$;

end

Calculate the normalized weights $W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{j=1}^{N} w_t^{(j)}}$ *(Correction step).*

end

Result: Weighted random samples $\left\{\mathbf{x}_{1:t}^{(j)},\ W_t^{(j)}\right\}_{j=1}^{N}$ approximating $\widetilde{\pi}_t$, for all $t = 1, \ldots, T$;

This method can be seen as a sequence of IS steps, where the target distribution at each step t is $\widetilde{\gamma}_t(\mathbf{x}_{1:t})$ (the unnormalized version of $\widetilde{\pi}_t$) and the importance distribution is given by

$$\widetilde{q}_t(\mathbf{x}_{1:t}) = \widetilde{q}_1(\mathbf{x}_1) \prod_{j=2}^{t} K_j(\mathbf{x}_{j-1}, \mathbf{x}_j), \tag{3.6}$$

where $K_j(\mathbf{x}_{j-1}, \,\cdot\,)$ is the mechanism used to propagate particles from time $t-1$ to t, known as the mutation stage. The algorithm works in the following way:

If $\left\{\mathbf{x}_{1:t}^{(j)},\ W_t^{(j)}\right\}_{j=1}^{N}$ is a set of weighted particles returned by the SMC algorithm then

$$\sum_{j=1}^{N} W_t^{(j)} \varphi(\mathbf{x}_{1:t}^{(j)}) \longrightarrow \mathbb{E}_{\widetilde{\pi}_t}[\varphi(\mathbf{X}_{1:t})] = \int_{E^t} \varphi(\mathbf{x}_{1:t}) \widetilde{\pi}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t}, \tag{3.7}$$

$\widetilde{\pi}_t$–almost surely as $N \to +\infty$, for any test function φ such that the expectation of φ under $\widetilde{\pi}_t$ exists.

Remark 3.4.1. *The reader should note that the knowledge of $\widetilde{\pi}_t$ up to a normalizing constant is sufficient for the implementation of a generic SMC algorithm, since the normalized version of the weights would be the same for both $\widetilde{\pi}_t$ and $\widetilde{\gamma}_t$, as also discussed in 3.2.*

The optimal selection of the mutation kernel (SMC importance distribution) for SMC methods is widely studied and a good tutorial review on the optimal choice minimizing the variance

of the incremental importance sampling weights is overviewed in [Doucet and Johansen, 2009]. There are also a range of known probabilistic properties of the SMC algorithm available in the literature, for a tutorial in the insurance context on these properties see [Del Moral et al., 2013]. This includes details on central limit theorem results for SMC algorithms along with asymptotic variance expressions, finite sample bias decompositions and propagation of chaos as well as finite sample concentration inequality bounds. There are also tutorials available on SMC algorithms in general such as [Doucet and Johansen, 2009] and the book length coverage of [Del Moral, 2004].

3.4.1 Resampling and moving particles

In practice the generic algorithm presented in the previous section will eventually (as t increases) be based only in a few distinct particles, in the sense that almost all the other particles will have negligible weights, in a phenomenon is known as *particle degeneracy*.

To overcome this problem, when the system is too degenerate one can resample all the particles $x_{1:t}$ after the correction step, choosing the j-th one with probability proportional to $W_t^{(j)}$. In [Liu and Chen, 1998] it was suggested using the Effective Sample Size (ESS) to measure the sample degeneracy, where its value at time t is defined as

$$ESS_t = \left[\sum_{j=1}^{N} (W_t^{(j)})^2\right]^{-1}$$

and resample steps should be performed only when $ESS_t < M$ – as a rule of thumb we can set $M = N/2$. It is important to note that after this step we need to set $W_t^{(j)} = 1/N$ for all particles, since they are all identically distributed.

Although the resample step alleviates the degeneracy problem, its successive reapplication, at each stage of the sampler, produces the so-called sample impoverishment, where the number of distinct particles is extremely small. In order to regenerate the system [Gilks and Berzuini, 2001] proposed to add a move with any kernel such that the target distribution is invariant with respect to it to rejuvenate the system. This kernel may be, for example, a Markov Chain kernel, which would begin with equally weighted samples from the target distribution and then perturb them under a single step of a Metropolis Hastings accept-reject mechanism. This would preserve the target distribution and add diversity to the particle cloud.

More precisely, we can apply any kernel $M(\mathbf{x}_{1:t}, \mathbf{x}^*_{1:t})$ that leaves $\widetilde{\pi}_t$ invariant to move particle $\mathbf{x}_{1:t}$ to $\mathbf{x}^*_{1:t}$ (the star will denote particles *after* the "move" step), i.e.,

$$\widetilde{\pi}_t(\mathbf{x}^*_{1:t}) = \int M(\mathbf{x}_{1:t}, \mathbf{x}^*_{1:t}) \widetilde{\pi}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t}.$$

Two of the simplest ways to construct such a kernel M are to use a Gibbs sampler or a Metropolis-Hastings (MH) algorithm (see Section 3.3). To use a Gibbs sampler algorithm, the full conditional distributions

$\widetilde{\pi}_t(\mathbf{x}_{1:t,i} \,|\, \mathbf{x}_{1:t,1}, \ldots, \mathbf{x}_{1:t,i-1}, \mathbf{x}_{1:t,i+1}, \ldots, \mathbf{x}_{1:t,d})$, for $i = 1, \ldots, d$, must be known up to proportionality, while for the MH they are not necessary. On the other hand, in the M-H algorithm one needs to design a proposal density $Q(\mathbf{x}_{1:t}, \mathbf{x}^*_{1:t})$ that moves the particle $\mathbf{x}_{1:t}$ to $\mathbf{x}^*_{1:t}$ or some

component of it such as $\mathbf{x}_t$ to $\mathbf{x}_t^*$. In the present context, the Gibbs sampler is described in Algorithm 7.

Algorithm 7: Gibbs Sampler algorithm.

Inputs: Full conditional p.d.f.'s:

$\widetilde{\pi}_t(\mathbf{x}_{1:t,1} \,|\, \mathbf{x}_{1:t,2}, \ldots, \mathbf{x}_{1:t,d}), \ldots, \widetilde{\pi}_t(\mathbf{x}_{1:t,d} \,|\, \mathbf{x}_{1:t,1}, \ldots, \mathbf{x}_{1:t,d})$; Sample from $\widetilde{\pi}_t$:

$\mathbf{x}_{1:t} = (\mathbf{x}_{1:t,1}, \ldots, \mathbf{x}_{1:t,d})$;

Sample $\mathbf{x}_{1:t,1}^* \sim \widetilde{\pi}_t(\mathbf{x}_{1:t,1} \,|\, \mathbf{x}_{1:t,2}, \ldots, \mathbf{x}_{1:t,d})$;

Sample $\mathbf{x}_{1:t,2}^* \sim \widetilde{\pi}_t(\mathbf{x}_{1:t,2} \,|\, \mathbf{x}_{1:t,1}^*, \mathbf{x}_{1:t,3}, \ldots, \mathbf{x}_{1:t,d})$;

$\vdots$

Sample $\mathbf{x}_{1:t,d}^* \sim \widetilde{\pi}_t(\mathbf{x}_{1:t,d} \,|\, \mathbf{x}_{1:t,1}^*, \ldots, \mathbf{x}_{1:t,d-1}^*)$;

Result: New sample from $\widetilde{\pi}_t$: $\mathbf{x}_{1:t}^* = (\mathbf{x}_{1:t,1}^*, \ldots, \mathbf{x}_{1:t,d}^*)$;

Including both the resampling and the "move" steps into the generic SMC algorithm leads to the "Resample-Move" algorithm, first presented in [Gilks and Berzuini, 2001] and subsequently widely used in the SMC literature.

The generic class of SMC algorithms whilst widely used in practice cannot be directly applied to the problems addressed in this work, since all the distributions in the sequence (4.1) are defined over the same support, i.e., $E_t = E$ and not $E_t = E \times \ldots \times E$ as required by the SMC algorithms just described. To overcome this problem a specialized variation of this method, named SMC Samplers is introduced in the next section but before we briefly revise some techniques used to estimate the normalizing constant using SMC methods.

3.4.2 Estimating the normalizing constant

In some situations the interest is not only to sample from the distribution $\widetilde{\pi}_t$, but also to compute its normalization constant $\mathcal{Z}_t$, which can be, for example, the model evidence (used to calculate Bayes factors) or the probability of a rare event.

At the t-th iteration of the SMC algorithm the normalizing constant $\mathcal{Z}_t$ can be written as

$$\begin{aligned}\mathcal{Z}_t &= \int_{E^t} \widetilde{\gamma}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t} \\ &= \int_{E^t} \frac{\widetilde{\gamma}_t(\mathbf{x}_{1:t})}{\widetilde{q}_t(\mathbf{x}_{1:t})} \widetilde{q}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t} \\ &= \int_{E^t} w_t \widetilde{q}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t}\end{aligned}$$

which, using the set of *unnormalized* samples $\left\{\mathbf{x}_{1:t}^{(j)},\ w_t^{(j)}\right\}_{j=1}^N$ can be approximated as

$$\mathcal{Z}_t = \int_{E^t} w_t \widetilde{q}_t(\mathbf{x}_{1:t}) d\mathbf{x}_{1:t} \approx \frac{1}{N} \sum_{j=1}^N w_t^{(j)} = \widehat{\mathcal{Z}}_t.$$

Defining $w_0 = 1$ the estimator shown above can be rewritten as

$$\widehat{\mathcal{Z}}_t = \frac{1}{N} \sum_{j=1}^N w_t^{(j)} = \frac{1}{N} \sum_{j=1}^N w_{t-1}^{(j)} \widetilde{\alpha}_t^{(j)} = \frac{1}{N} \sum_{j=1}^N \prod_{k=1}^t \widetilde{\alpha}_t^{(j)}, \tag{3.8}$$

where $\widetilde{\alpha}_t^{(j)}$ is the incremental weight defined in Algorithm 6.

Moreover, defining $\widehat{\mathcal{Z}}_0 = 1$ we find the often presented form for the estimator of the normalizing constant, namely

$$\begin{aligned}
\widehat{\mathcal{Z}}_t &= \frac{\widehat{\mathcal{Z}}_1}{\widehat{\mathcal{Z}}_0} \cdots \frac{\widehat{\mathcal{Z}}_d}{\widehat{\mathcal{Z}}_{d-1}} \\
&= \prod_{k=1}^{t} \frac{\widehat{\mathcal{Z}}_k}{\widehat{\mathcal{Z}}_{k-1}} \\
&= \prod_{k=1}^{t} \frac{\sum_{j=1}^{N} w_k^{(j)}}{\sum_{i=1}^{N} w_{k-1}^{(i)}} \\
&= \prod_{k=1}^{t} \sum_{j=1}^{N} \frac{w_{k-1}^{(j)}}{\sum_{i=1}^{N} w_{k-1}^{(i)}} \frac{w_k^{(j)}}{w_{k-1}^{(j)}} \\
&= \prod_{k=1}^{t} \sum_{j=1}^{N} W_{k-1}^{(j)} \widetilde{\alpha}_k^{(j)} \\
&= \widehat{\mathcal{Z}}_{t-1} \sum_{j=1}^{N} W_{t-1}^{(j)} \widetilde{\alpha}_t^{(j)},
\end{aligned}$$

where $W_{k-1}^{(j)}$ is the normalized weight defined in Algorithm 6.

Remark 3.4.2. *The estimator defined in (3.8) can be proved to be unbiased and asymptotically normally distributed when the number of particles $N \to +\infty$ (see [Del Moral, 2004, Propositions 7.4.1 and 9.4.1] and [Pitt et al., 2012] for a proof in the special case of state-space models).*

Other estimators for (ratios of) normalizing constants have been proposed in the IS/SMC literature, such as the path sampling of [Gelman and Meng, 1998] used in a SMC for rare event set up in [Johansen et al., 2006]. This estimator is defined for a continuous *path* of distributions denote as

$$\pi_{\theta(t)}(\mathbf{x}) = \mathcal{Z}^{-1} \gamma_{\theta(t)}(\mathbf{x}),$$

with $t \in [0, 1]$, $\theta(0) = 0$ and $\theta(1) = 1$. The estimator is, then, based on the following (path sampling) identity

$$\log\left(\frac{\mathcal{Z}_1}{\mathcal{Z}_\prime}\right) = \int_0^1 \frac{d\theta(t)}{dt} \int \frac{d\log(\gamma_{\theta(t)}(\mathbf{x}))}{dt} \pi_{\theta(t)}(\mathbf{x}) d\mathbf{x} dt.$$

For the algorithms proposed to compute risk allocations the path sampling approach is not applicable, as the estimator relies on derivatives of the schedule (i.e., $\theta(t)$) with respect to t which, in our framework, are not readily available. Instead, we discuss a problem-specific algorithm in Section 6.4.1.

3.4.3 SMC samplers

Having presented the SMC class of algorithms, we now present in contrast to these the class of SMC Sampler algorithms which involve the same mechanism as the SMC algorithm also using a mutation, correction and resampling stage at each iteration. However, the class of SMC Sampler algorithm is importantly different in the space that the sequence of distributions being sampled from are defined upon. Differently from Section 3.4, our interest now is to approximate

a generic sequence of probability distributions $\{\pi_t\}_{t=1}^T$ such that $supp(\pi_t) = supp(\pi_{t-1}) = E$ for all $t = 1, \ldots, T$ (once again one may think of $E = \mathbb{R}^d$). Here we may also assume our target distribution is only known up to a normalizing constant, i.e., $\pi(\mathbf{x}) = \mathcal{Z}^{-1}\gamma(\mathbf{x})$.

Remark 3.4.3. *For notational clarity, functions in the enlarged space will be denoted with a upper tilde, as* $\widetilde{\gamma}_t : E^t \longrightarrow \mathbb{R}$.

The idea presented on [Peters, 2005] and [Del Moral et al., 2006] is to transform this problem into one resembling the usual SMC formulation, where the sequence of target distributions $\{\widetilde{\gamma}_t\}_{t=1}^T$ is defined on the product space, i.e., $supp(\widetilde{\gamma}_t) = E \times E \times \ldots \times E = E^t$.

The construction of $\widetilde{\gamma}_t$ (the density in the *path* space) is carried out as:

$$\widetilde{\pi}_t(\mathbf{x}_{1:t}) \propto \widetilde{\gamma}_t(\mathbf{x}_{1:t}) = \gamma_t(\mathbf{x}_t)\widetilde{\gamma}_t(\mathbf{x}_{1:t-1}|\mathbf{x}_t), \text{ for } t = 2, \ldots, T \tag{3.9}$$

where $\widetilde{\gamma}_t(\mathbf{x}_{1:t-1}|\mathbf{x}_t)$ is a probability distribution on the space E^{t-1}, for all $\mathbf{x}_t \in E$. Similarly to (3.6) we can carry out the construction of the importance distribution at time t.

As noticed in [Peters, 2005] and [Del Moral et al., 2006], a wise choice for $\widetilde{\gamma}_t(\mathbf{x}_{1:t-1}|\mathbf{x}_t)$ is given by

$$\widetilde{\gamma}_t(\mathbf{x}_{1:t-1}|\mathbf{x}_t) = \prod_{s=1}^{t-1} L_s(\mathbf{x}_{s+1}, \mathbf{x}_s),$$

where the each L_s is the density of an (artificial) backward Markov kernel. It is important to note that, by construction, $\widetilde{\gamma}_t(\mathbf{x}_{1:t})$ admits $\gamma_t(\mathbf{x}_t)$ as a marginal, since

$$\int \widetilde{\gamma}_t(\mathbf{x}_{1:t})d\mathbf{x}_{1:t-1} = \gamma_t(\mathbf{x}_t) \int \prod_{s=1}^{t-1} L_s(\mathbf{x}_{s+1}, \mathbf{x}_s)d\mathbf{x}_{1:t-1} = \gamma_t(\mathbf{x}_t), \ \forall t > 1.$$

Moreover, provided that $\widetilde{\gamma}_t$ admits γ_t as a marginal, the normalizing constant of the enlarged density will be the same as the original density:

$$\int \widetilde{\gamma}_t(\mathbf{x}_{1:t})d\mathbf{x}_{1:t} = \int\int \widetilde{\gamma}_t(\mathbf{x}_{1:t})d\mathbf{x}_{1:t-1}d\mathbf{x}_t = \int \gamma_t(\mathbf{x}_t)d\mathbf{x}_t = \mathcal{Z}_t.$$

Now that we are back to the SMC framework from last section, we can easily write the SMC Sampler algorithm (Algorithm 8). Moreover, the Resample-Move strategy from Section 3.4.1 can still be utilized.

3.4.3.1 Backward kernels selection

The introduction of the sequence of kernels $\{L_{t-1}\}_{t=2}^T$ creates a new degree of freedom in SMC samplers when compared with usual SMC algorithms, where only the forward mutation kernels $\{K_t\}_{t=1}^T$ should be designed. In this section we will discuss how to, given the kernels $\{K_t\}_{t=1}^T$, optimize the choice of backward kernels $\{L_{t-1}\}_{t=2}^T$.

Denote by $q_t(\mathbf{x}_t)$ the marginal importance distribution at time t, which is given by

$$q_t(\mathbf{x}_t) = \int \widetilde{q}_t(\mathbf{x}_{1:t})d\mathbf{x}_{1:t-1} = \int q_1(\mathbf{x}_1) \prod_{j=2}^{t} K_j(\mathbf{x}_{j-1}, \mathbf{x}_j)d\mathbf{x}_{1:t-1}. \tag{3.10}$$

In the case in which we know how to calculate q_t in exact form we can simply approximate the target distribution γ_t by a weighted sample $\{\mathbf{x}_t^{(j)}, W_t^{(j)}\}$, where $\mathbf{x}_t \sim q_t$ and W_t is the

Algorithm 8: SMC Sampler algorithm.

Inputs: IS density q_1, (forward) mutation kernels $\left\{K_t(\mathbf{x}_{t-1}, \mathbf{x}_t)\right\}_{t=1}^{T}$, (artificial) backward kernels $\left\{L_{t-1}(\mathbf{x}_t, \mathbf{x}_{t-1})\right\}_{t=1}^{T}$, move kernel $M_t(\widehat{\mathbf{x}}_t, \mathbf{x}_t)$;

for $j = 1, \ldots, N$ **do**

Sample $\mathbf{x}_1^{(j)}$ from $\mathbf{X}_1 \sim \widetilde{q}_1(\,\cdot\,)$ *(Mutation step)*;

Calculate the weights $w_1^{(j)} = \frac{\widetilde{\gamma}_1(\mathbf{x}_1^{(j)})}{\widetilde{q}_1(\mathbf{x}_1^{(j)})}$;

end

Calculate the normalized weights $W_1^{(j)} = \frac{w_1^{(j)}}{\sum_{j=1}^{N} w_1^{(j)}}$ *(Correction step)*;

for $t = 2, \ldots, T$ **do**

for $j = 1, \ldots, N$ **do**

Sample $\mathbf{x}_t^{(j)}$ from $\mathbf{X}_t \big| \mathbf{X}_{t-1} = \mathbf{x}_{t-1}^{(j)} \sim K_t(\mathbf{x}_{t-1}^{(j)}, \,\cdot\,)$ *(Mutation step)*;

Using (3.9) and (3.6), calculate the weights

$$w_t^{(j)} = \frac{\widetilde{\gamma}_t(\mathbf{x}_{1:t}^{(j)})}{\widetilde{q}_t(\mathbf{x}_{1:t}^{(j)})} = w_{t-1}^{(j)} \underbrace{\frac{\gamma_t(\mathbf{x}_t^{(j)}) L_{t-1}(\mathbf{x}_t^{(j)}, \mathbf{x}_{t-1}^{(j)})}{\gamma_{t-1}(\mathbf{x}_{t-1}^{(j)}) K_t(\mathbf{x}_{t-1}^{(j)}, \mathbf{x}_t^{(j)})}}_{\text{incremental weight: } \alpha(\mathbf{x}_{t-1}^{(j)}, \mathbf{x}_t^{(j)})} \;;$$

end

Calculate the normalized weights $W_t^{(j)} = \frac{w_t^{(j)}}{\sum_{j=1}^{N} w_t^{(j)}}$ *(Correction step).* ;

if $ESS_t < N/2$ **then**

for $j = 1, \ldots, N$ **do**

Resample $\widehat{\mathbf{x}}_t^{(j)} = \mathbf{x}_t^{(k)}$ with prob. $W_t^{(k)}$ *(Resample step)* ;

Sample $\mathbf{x}_t^{(j)} \sim M_t(\widehat{\mathbf{x}}_t^{(j)}, \,\cdot\,)$ *(Move step)*;

Set $W_t^{(j)} = 1/N$;

end

end

end

Result: Weighted random samples $\left\{\mathbf{x}_t^{(j)},\ W_t^{(j)}\right\}_{j=1}^{N}$ approximating π_t, for all $t = 1, \ldots, T$;

normalized version of $w_t = \frac{\gamma_t(\mathbf{x}_t)}{q_t(\mathbf{x}_t)}$. From the definition of q_t we can see that sampling from q_t is simple if it is easy to sample from q_1 and from all the kernels K_t. On the other hand, the density of q_t will only be tractable if we are able to solve the marginalization integral (in $t-1$ dimensions) – which, in practice will hardly ever be the case.

The introduction of backward kernels $\{L_{t-1}\}_{t=2}^T$ helps us (in most of the practical cases) to avoid the computation of q_t. On the other hand, since $\widetilde{\gamma}_t$ and $\widetilde{q}_t$ admits, respectively, γ_t and q_t as marginals, Lemma 3.4.4 tells us the price we need to pay: *an increase in the variance of the importance weights*. Fortunately, the same Lemma provides us some insights on how to optimally choose the backward kernels.

Lemma 3.4.4. *Let $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and $g(\boldsymbol{x}_1, \boldsymbol{x}_2)$ be two probability densities with $supp(f) \subset supp(g)$. Then*

$$Var_g\left(\frac{f(\boldsymbol{X}_1, \boldsymbol{X}_2)}{g(\boldsymbol{X}_1, \boldsymbol{X}_2)}\right) \geq Var_g\left(\frac{\gamma_1(\boldsymbol{X}_1)}{g_1(\boldsymbol{X}_1)}\right),$$

where $f_1(\boldsymbol{x}_1) = \int f(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{x}_2$ and $g_1(\boldsymbol{x}_1) = \int g(\boldsymbol{x}_1, \boldsymbol{x}_2) d\boldsymbol{x}_2$.

Proof. From the variance decomposition, we have that

$$\begin{aligned}
\mathrm{Var}_g\left(\frac{f(\mathbf{X}_1, \mathbf{X}_2)}{g(\mathbf{X}_1, \mathbf{X}_2)}\right) &= \mathrm{Var}_g\left(\mathbb{E}_g\left[\frac{f(\mathbf{X}_1, \mathbf{X}_2)}{g(\mathbf{X}_1, \mathbf{X}_2)}\Big|\mathbf{X}_1 = \mathbf{x}_1\right]\right) \\
&\quad + \mathbb{E}_g\left[\mathrm{Var}_g\left(\frac{f(\mathbf{X}_1, \mathbf{X}_2)}{g(\mathbf{X}_1, \mathbf{X}_2)}\Big|\mathbf{X}_1 = \mathbf{x}_1\right)\right] \\
&\geq \mathrm{Var}_g\left(\mathbb{E}_g\left[\frac{f(\mathbf{X}_1, \mathbf{X}_2)}{g(\mathbf{X}_1, \mathbf{X}_2)}\Big|\mathbf{X}_1 = \mathbf{x}_1\right]\right),
\end{aligned}$$

since $f, g \geq 0$ (they are densities).

The result follows from the fact that the ratio of marginal densities can be rewritten as the following conditional expectation:

$$\begin{aligned}
\frac{f_1(\mathbf{x}_1)}{g_1(\mathbf{x}_1)} &= \int \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{g_1(\mathbf{x}_1) g_{2|1}(\mathbf{x}_2|\mathbf{x}_1)} g_{2|1}(\mathbf{x}_2|\mathbf{x}_1) d\mathbf{x}_2 \\
&= \mathbb{E}_g\left(\frac{f(\mathbf{X}_1, \mathbf{X}_2)}{g(\mathbf{X}_1, \mathbf{X}_2)}\Big|\mathbf{X}_1 = \mathbf{x}_1\right).
\end{aligned}$$

□

As mentioned previously, Proposition 3.4.5 shows how to design the backward kernels $\{L_{t-1}\}_{t=2}^T$ in order to minimize the variance of the importance weights.

Proposition 3.4.5 (Optimal backward kernel)**.** *The kernel $L_t^{opt}(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t) = \dfrac{q_t(\boldsymbol{x}_t) K_{t+1}(\boldsymbol{x}_t, \boldsymbol{x}_{t+1})}{q_{t+1}(\boldsymbol{x}_{t+1})}$ is optimal in the sense that $Var_{q_t}^{\sim}(w_t^{opt}(\boldsymbol{X}_{1:t})) \leq Var_{q_t}^{\sim}(w_t(\boldsymbol{X}_{1:t}))$, where $w_t^{opt}(\boldsymbol{x}_{1:t}) = \dfrac{\gamma_t(\boldsymbol{x}_t)}{q_t(\boldsymbol{x}_t)}$.*

Proof. If we substitute the optimal kernel into the definition of importance weights we have that

$$\begin{aligned} w_t^{opt}(\mathbf{x}_{1:t}) &= \frac{\widetilde{\gamma}_t(\mathbf{x}_{1:t})}{\widetilde{q}_t(\mathbf{x}_{1:t})} \\ &= \frac{\gamma_t(\mathbf{x}_t)\prod_{s=1}^{t-1} L_s^{opt}(\mathbf{x}_{s+1},\mathbf{x}_s)}{q_1(\mathbf{x}_1)\prod_{j=2}^{t} K_j(\mathbf{x}_{j-1},\mathbf{x}_j)} \\ &= \frac{\gamma_t(\mathbf{x}_t)\prod_{s=1}^{t-1} \frac{q_t(\mathbf{x}_t)K_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})}{q_{t+1}(\mathbf{x}_{t+1})}}{q_1(\mathbf{x}_1)\prod_{j=2}^{t} K_j(\mathbf{x}_{j-1},\mathbf{x}_j)} \\ &= \frac{\gamma_t(\mathbf{x}_t)}{q_t(\mathbf{x}_t)}. \end{aligned}$$

The result, then, follows from Lemma 3.4.4.

□

From Proposition 3.4.5 we can see that if we know how to sample from $q_t(\mathbf{x}_t)$ then the SMC sampler algorithm reduces to a sequence of IS steps, where at each time t we sample particles from $q_t(\mathbf{x}_t)$ and correct the bias through the weights $w_t(\mathbf{x}_t) = \frac{\gamma_t(\mathbf{x}_t)}{q_t(\mathbf{x}_t)}$.

Independent kernel One situation where we know how to calculate $q_t(\mathbf{x}_t)$ is when $K_t(\mathbf{x}_{t-1},\mathbf{x}_t)$ does not depend on $\mathbf{x}_{t-1}$, making the mutation step completely memory-less. As an abuse of notation, let $K_t(\mathbf{x}_t) = K_t(\mathbf{x}_{t-1},\mathbf{x}_t)$. This choice is not always recommended to be used in practice due to difficulties in designing an appropriate kernel, but in this case it is easy to see that it is possible to perform a sequence of IS steps, since

$$\begin{aligned} q_t(\mathbf{x}_t) &= \int q_1(\mathbf{x}_1)\prod_{j=2}^{t} K_j(\mathbf{x}_{j-1},\mathbf{x}_j)d\mathbf{x}_{1:t-1} \\ &= \int q_1(\mathbf{x}_1)\prod_{j=2}^{t} K_j(\mathbf{x}_j)d\mathbf{x}_{1:t-1} \\ &= \left(\int q_1(\mathbf{x}_1)d\mathbf{x}_1\right)\left(\prod_{j=2}^{t-1}\int K_j(\mathbf{x}_j)d\mathbf{x}_{1:t-1}\right)\left(\int K_t(\mathbf{x}_t)d\mathbf{x}_t\right) \\ &= K_t(\mathbf{x}_t). \end{aligned}$$

Approximations of the optimal kernel Various approximations of the optimal backward kernel have been proposed in the literature (see, for example [Del Moral et al., 2006, Section 3.3.3]) but here we will discuss only one of them.

If we rewrite the optimal backward kernel from Proposition 3.4.5 as

$$L_t^{opt}(\mathbf{x}_{t+1},\mathbf{x}_t) = \frac{q_t(\mathbf{x}_t)K_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})}{\int q_t(\mathbf{x}_t)K_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})d\mathbf{x}_t}$$

it suggests that a sensible approximation for this kernel is to use π_t instead of q_t. In this case,

$$L_t^{opt}(\mathbf{x}_{t+1},\mathbf{x}_t) \approx \frac{\gamma_t(\mathbf{x}_t)K_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})}{\int \gamma_t(\mathbf{x}_t)K_{t+1}(\mathbf{x}_t,\mathbf{x}_{t+1})d\mathbf{x}_t}, \tag{3.11}$$

since the normalizing constants of π cancel out.

Although the integral in the denominator of (3.11) is usually not analytically tractable we can use the weighted sample $\{\mathbf{x}_{1:t}^{(j)}, w_t^{(j)}\}_{j=1}^N$ from π_t generated by the SMC sampler procedure

to approximate L_t^{opt} by

$$L_t(\mathbf{x}_{t+1}, \mathbf{x}_t) = \frac{\gamma_t(\mathbf{x}_t) K_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1})}{\frac{1}{N}\sum_{j=1}^N w_t^{(j)} K_{t+1}(\mathbf{x}_t^{(j)}, \mathbf{x}_{t+1})}, \tag{3.12}$$

which leads to incremental weights

$$\alpha(\mathbf{x}_{t-1}, \mathbf{x}_t) = \frac{\gamma_t(\mathbf{x}_t)}{\frac{1}{N}\sum_{j=1}^N w_{t-1}^{(j)} K_t(\mathbf{x}_{t-1}^{(j)}, \mathbf{x}_t)}. \tag{3.13}$$

3.4.3.2 Forward kernels selection

So far we have discussed how to design backward kernels L_t that are optimal for specific choices of forward kernels K_t. We now present possible choices of K_t in order to have simple forms of importance weights when used with the optimal choice of backward kernel.

In this sense, one convenient choice for the forward kernels K_t is to assume they are such that $K_t(\mathbf{x}_{t-1}, \mathbf{x}_t)$ has π_t as invariant density, i.e.,

$$\pi_t(\mathbf{x}_{t+1}) = \int K_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}) \pi_{t+1}(\mathbf{x}_t) d\mathbf{x}_t.$$

Proceeding this way, we can choose the backward kernel as follows,

$$L_t(\mathbf{x}_{t+1}, \mathbf{x}_t) = \frac{\pi_{t+1}(\mathbf{x}_t) K_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1})}{\pi_{t+1}(\mathbf{x}_{t+1})}, \tag{3.14}$$

which is a reasonable approximation for the optimal backward kernel from Proposition 3.4.5. It also worth noticing the kernel defined on (3.14) is the reversal Markov kernel associated with K_{t+1}.

For theoretical purposes, we may also assume the forward kernel mixes perfectly, i.e., $K_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}) = \pi_{t+1}(\mathbf{x}_{t+1})$. This choice of kernels is obviously not feasible in practice, since $\pi_{t+1}(\mathbf{x}_{t+1})$ is precisely the density we are trying to sample from, but it can provide interesting insights. In this case, the incremental weights of the SMC sampler algorithm (see Algorithm 8) are given by

$$\alpha(\mathbf{x}_{t-1}, \mathbf{x}_t) = \frac{\gamma_t(\mathbf{x}_t) L_{t-1}(\mathbf{x}_t, \mathbf{x}_{t-1})}{\gamma_{t-1}(\mathbf{x}_{t-1}) K_t(\mathbf{x}_{t-1}, \mathbf{x}_t)} \propto \frac{\pi_t(\mathbf{x}_{t-1})}{\pi_{t-1}(\mathbf{x}_{t-1})},$$

which makes the weights at time t independent of the particles sampled at time t.

3.5 Simulation methods in the copula space

In this section we introduce the idea of transforming the original risk/capital allocation problem from sampling a $\mathbb{R}^d$-valued constrained random variable to sampling from a constrained copula, in an algorithm to be made precise in Chapter 4.

As seen in equation (3.1), in order to calculate risk/capital allocations we need to compute expectations of the form $\mathbb{E}[h(\mathbf{X}) \,|\, g(\mathbf{X}) \in A]$. From the discussion in the previous sections, this expectation can be approximated via a Monte Carlo simulation through the use of a set of N *weighted samples* $\{\mathbf{x}^{(j)}, W^{(j)}\}_{j=1}^N$ from the conditional distribution $f_{\mathbf{X}}(\mathbf{x} \,|\, g(\mathbf{x}) \in A)$, where $\sum_{j=1}^N W^{(j)} = 1$. The approximation is, then, given by

$$\mathbb{E}[h(\mathbf{X}) \,|\, g(\mathbf{X}) \in A] \approx \sum_{j=1}^N W^{(j)} h(\mathbf{x}^{(j)}).$$

For example, if we can directly sample i.i.d. realizations from the distribution of $\mathbf{X} \,|\, g(\mathbf{X}) \in A$ (e.g., using a rejection scheme) we would have $W^{(j)} = 1/N$. In general, though, samples from the above conditional distribution are not readily available and we need to use an importance sampling distribution (coupled with a rejection step), calculating the weights accordingly to remove bias. However, if done naively this would result in very large rejection rates that would behave poorly as the conditioning event becomes rarer. In the sequel we discuss a natural transformation of multivariate random variables, which allows us to perform the sampling procedure in the d-dimensional unit cube $[0,1]^d$.

For the expectations we are interested in, the conditioning event can be defined as $[\mathbf{X} \in \mathcal{G}_{\mathbf{X}}]$, where

$$\mathcal{G}_{\mathbf{X}} := \{\mathbf{x} \in \mathbb{R}^d \,:\, g(\mathbf{x}) \in A\}, \tag{3.15}$$

since $g(\mathbf{X}) \in A \iff \mathbf{X} \in \mathcal{G}_{\mathbf{X}}$. Given that our multivariate loss model is uniquely characterized by a copula (either explicitly or implicitly, through a parametric joint distribution) the region $\mathcal{G}_{\mathbf{X}}$ in $\mathbb{R}^d$ holds a close relationship with some region in $[0,1]^d$. Formally, if we define

$$\mathcal{G}_{\mathbf{U}} := \left\{\mathbf{u} \in [0,1]^d \,:\, \left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \in \mathcal{G}_{\mathbf{X}}\right\} \tag{3.16}$$

then it holds that

$$\mathbf{x} \in \mathcal{G}_{\mathbf{X}} \iff \mathbf{u} \in \mathcal{G}_{\mathbf{U}}.$$

Therefore, similarly to the simulation of an unconditional multivariate distribution, to sample $\mathbf{x} = (x_1, \ldots, x_d)$ from the distribution of $\Big(\mathbf{X} \,|\, g(\mathbf{X}) \in A\Big)$ we can

1. Produce a weighted sample $\{\mathbf{u}^{(j)}, W^{(j)}\}_{j=1}^N$ from C such that $\mathbf{u}^{(j)} = (u_1^{(j)}, \ldots, u_d^{(j)}) \in \mathcal{G}_{\mathbf{U}}$ for all $j = 1, \ldots, N$;

2. Return the weighted sample $\{\mathbf{x}^{(j)}, W^{(j)}\}_{j=1}^N$ where $x_i^{(j)} = F_i^{-1}(u_i^{(j)})$, for $i = 1, \ldots, d$, $j = 1, \ldots, N$.

Note that one can calculate conditional expectations with respect to $\mathbf{X}$ as follows

$$\mathbb{E}[h(\mathbf{X}) \,|\, g(\mathbf{X}) \in A] = \mathbb{E}\Big[h\big(F_1^{(-1)}(U_1), \ldots, F_d^{(-1)}(U_d)\big) \,|\, \mathbf{U} \in \mathcal{G}_{\mathbf{U}}\Big] \tag{3.17}$$

$$\approx \sum_{j=1}^{N} W^{(j)} h\big(F_1^{(-1)}(U_1^{(j)}), \ldots, F_d^{(-1)}(U_d^{(j)})\big). \tag{3.18}$$

Clearly, if all the marginal quantile functions F_i^{-1} are known, then the difficulty of the proposed approach is to sample from the constrained copula. The idea of performing the sampling procedure in the constrained copula space has been independently developed by Arbenz, Cambou and Hofert in [Arbenz et al., 2014] (see an overview of their algorithm in Section 3.6.1), where an importance sampling distribution is designed to target the distribution of $\mathbf{u} \,|\, \mathbf{u} \in \mathcal{G}_{\mathbf{U}}$.

In the algorithm developed in Chapter 4, instead of targeting the rare region $\mathbf{u} \in \mathcal{G}_{\mathbf{U}}$ we propose to sequentially target less rare regions, in a specially designed SMC Sampler procedure that is made precise in Sections 4.1 and 3.4.

Before presenting these specialized algorithms, it is informative to briefly comment on alternative Importance Sampling (IS) based approaches in the actuarial literature. At this stage we observe that there are many different types of rare-event simulation algorithms available, and the choice of a particular type will depend principally on how one defines the notion of a "rare-event" in the sample space. Although the following brief discussions on alternative IS based solutions are not directly targeting the same type of multivariate rare-event problems as faced in the case of capital allocation, they are informative to discuss especially with regard to the concept of relative error.

3.6 Related simulation approaches for rare events

Differently from the problem we consider, where the interest lies on calculating expectations conditional to a rare event, most of the literature on rare event simulation focus on the computation of the probability of a rare event, a quantity that is perfectly known in our framework. Nonetheless, some of these methods hold a strong relationship with our proposed SMC algorithm, in particular the class of *splitting* methods.

In the context of rare event estimation one of the most widespread methodology relates to the broad field of *splitting* algorithms, which dates as far back as the 1950's (see [Kahn and Harris, 1951]). In order to calculate the probability of a stochastic process reaching a rare event the authors propose to simulate many trajectories and duplicate those that approach the event of interest, splitting the probability of interest in the product of easier to compute terms.

This idea was rediscovered several years later in [Villén-Altamirano and Villén-Altamirano, 1991] and named RESTART (REpetitive Simulation Trials After Reaching Thresholds) and since then it has been considerably studied and improved, see, for example, [Glasserman et al., 1996], [Glasserman et al., 1999], [Garvels, 2000], [Au and Beck, 2001] (under the name of *subset simulation*) and [L'Ecuyer et al., 2006].

These algorithms were linked with SMC methods and rigorously studied in [Cérou et al., 2005], [Cérou et al., 2006], [Cérou and Guyader, 2007] and, more recently, [Cérou et al., 2012].

The idea of introducing intermediate steps in order to estimate the probability of rare events has also been recently discussed in the new MCMC-based *split sampling* algorithm of [Birge et al., 2012], related to the *nested sampling* algorithm of [Skilling, 2006].

Other simulation based algorithms for rare event simulation include the Esscher transform, introduced in the actuarial literature in [Escher, 1932] and used, for example, in [McLeish, 2010]; and the cross entropy method (see [Kroese et al., 2013]).

Related to the allocation problem studied here, [Glasserman, 2005] presents, in the particular case of Gaussian copulas, an IS scheme to approximate conditional expectations. Also, for the same family of models, [Siller, 2013] more recently proposed a method based on Fourier transforms to compute marginal risk contributions.

We also note that there are classes of asymptotic approximation results available for ap-

proximation of capital allocations. For instance, in order to estimate expectations of the form (3.1) in a bivariate set-up, [Cai et al., 2015] assume that large values of $g(\mathbf{X})$ correspond to high values of $h(\mathbf{X})$ (in their case $h(\mathbf{X}) = X_i$). Under these constraints, the authors use results from Extreme Value Theory (EVT) to derive an estimator of (3.1) and study some of its properties. We refer the interested reader to references in [Cai et al., 2015] for further background on such asymptotic approximations and instead we continue to focus upon sampling based solutions.

The closest class of IS based solution to our proposed SMC Sampler solution has been developed recently by [Arbenz et al., 2014]. We present this briefly before detailing our approach.

3.6.1 Arbenz-Cambou-Hofert algorithm

In common with the sampling method proposed in Chapter 4, the approach recently developed in [Arbenz et al., 2014] involves sampling from a target distribution given by a constrained copula. This is particularly relevant as it leads to convenient bounded state spaces for sampling, since the support of the copula when unconstrained is $[0, 1]^d$ and consequently the constrained copula will be on a sub-space of this hypercube.

Unlike our proposed solution, the approach of [Arbenz et al., 2014] involves developing an Importance Sampling (IS) scheme targeting the constrained copula distribution. However, in contrast to our approach their method does not involve any intermediate sequence of constrained regions leading smoothly up to the rare-event constraint. Instead, they try to directly approximate the optimal importance sampling proposal. This is in general a very challenging task and they have some interesting insight. For this reason we briefly present their methodology below as it will form a direct comparison with the approach we develop.

In [Arbenz et al., 2014] the aim of their work is to generate a sample from the unconditional copula with most of the particles satisfying a condition such as (3.16). In order to generate these samples, an importance sampling distribution $F_{\mathbf{V}}$ is designed as a mixture of conditional copulas. More formally, the IS distribution is defined as

$$F_{\mathbf{V}}(\mathbf{u}) = \int_0^1 C^{[\lambda]}(\mathbf{u}) dF_{\Lambda}(\lambda), \tag{3.19}$$

where $C^{[\lambda]}$ is the distribution of $\mathbf{U}$ conditional on the event that at least one of its components exceeds λ, i.e.,

$$C^{[\lambda]}(\mathbf{u}) = \mathbb{P}[U_1 \leq u_1, \ldots, U_d \leq u_d \,|\, \max\{U_1, \ldots, U_d\} > \lambda].$$

In the main algorithm presented in [Arbenz et al., 2014], samples from the importance distribution are generated by rejection, but a "conditional sampling algorithm" is also presented. Overall, an appealing aspect of their proposed method is that it does not make use of the copula density explicitly. This can be advantageous in settings in which the copula density is computationally expensive to be calculated or even unknown. However, as with all Monte Carlo methods, there are also drawbacks to the proposed approach that we will argue can be overcome through development not of an IS solution but instead via a SMC Sampler solution in the constrained copula space.

Under simplifying assumptions on the joint behaviour of $\mathbf{U}$, an optimal distribution for F_Λ is presented, but in general the only restriction on the choice of the mixing distribution F_Λ is that $\mathbb{P}[\Lambda = 0] > 0$. To sample from $\mathbf{X} \,|\, \mathbf{X} > B$ one of the algorithms proposed is given as follows.

Algorithm 9: IS-ACH algorithm from [Arbenz et al., 2014].

Inputs: N: desired sample size for $F_{\mathbf{V}}$; F_Λ: mixing distribution (as in (3.19)) ;

for $j = 1, \dots N$ **do**

Sample $\Lambda^{(j)} \sim F_\Lambda$;

Sample $\mathbf{u}^{(j)} \sim C$ until $\max\{u_1^{(j)}, \dots, u_d^{(j)}\} > \Lambda^{(j)}$;

Define $x_i^{(j)} = F_i^{-1}\left(u_i^{(j)}\right)$ for $i = 1, \dots, d$;

Compute the importance weight $w^{(j)} := w(\mathbf{u}^{(j)})$ as in [Arbenz et al., 2014], Section 5 ;

Compute the normalized importance weight $W^{(j)} = \frac{w^{(j)}}{\sum_{j=1}^{N} w^{(j)}}$;

Define $\{\widetilde{\mathbf{u}}^{(j_k)}\}_{k=1}^{\widetilde{N}}$ as the sub-set of $\{\mathbf{u}^{(j)}\}_{j=1}^{N}$ such that $\sum_{i=1}^{d} \widetilde{x}_i^{(j_k)} > B$;

end

Result: Weighted random samples (of random sample size $\widetilde{N}$): $\left\{\mathbf{u}^{(j)},\ W^{(j)}\right\}_{j=1}^{\widetilde{N}}$;

Some points need to be stressed about Algorithm 9. First, let $\mathbb{E}[N_{\mathbf{V}}]$ denote the expected number of draws from C in order to have a sample satisfying $\max\{u_1, \dots, u_d\} > \Lambda$. It can be easily shown that (see [Arbenz et al., 2014], Lemma 4.2)

$$\mathbb{E}[N_{\mathbf{V}}] = \int_0^1 \frac{1}{1 - C(\lambda \mathbf{1})} dF_\Lambda(\lambda),$$

where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^d$. Therefore, to generate a sample of fixed size N from $F_{\mathbf{V}}$ it is necessary to sample (on average) $N \times \mathbb{E}[N_{\mathbf{V}}]$ times from C.

Another important aspect of this algorithm pertains to an understanding of the number of "particles" (samples) which are obtained with non-zero weight. Since $p_0 = \mathbb{P}[\Lambda = 0]$ is necessarily positive, we can ensure some of the N samples from $F_{\mathbf{V}}$ will be actually from the unconditional copula C, meaning that $p_0 \times 0.99 \times 100\%$ of the particles are expected not to satisfy the condition $\sum_{i=1}^{d} X_i > \mathrm{VaR}_{0.99}(\sum_{i=1}^{d} X_i)$. For $\lambda > 0$ the same behaviour is expected, leading, in practice, to $\widetilde{N}$ (as defined in the last step of Algorithm 9) being smaller than N, and in cases of relevance to capital allocation, this difference can be significant, with $\widetilde{N} << N$. In capital allocation problems such cases can prove to be a serious problem in terms of computational cost and efficiency for this IS based approach as will be discussed in Section 4.3.

3.7 Final remarks

Note that in any practical scenario the quantile level α in the allocation problem will be chosen to be close to one and both conditioning events will have very small probabilities, i.e., $\mathbb{P}[g(\mathbf{X}) \in A] \approx 0$. More precisely, the event will have probability $1 - \alpha$ for the ES case and probability zero (regardless of the choice of α) in the VaR case, since the set A is a single point. Therefore in practice, in the VaR case one would work instead with an ϵ approximation by setting the conditioning event as the ϵ-ball of A given by $\mathcal{B}_\epsilon(A)$ specified according

to $\mathcal{B}_\epsilon(A) = \left[\mathrm{VaR}_\alpha(\sum_{i=1}^d X_i) - \epsilon,\, \mathrm{VaR}_\alpha(\sum_{i=1}^d X_i) + \epsilon\right]$ for some small positive ϵ in the neighbourhood of zero. This type of approach was advocated in [Glasserman, 2005] and also in [Del Moral et al., 2013].

From a risk management perspective, many other choices of h and g can be of interest. For example, if one is interested in measuring the impact of marginal tail events in the portfolio, one could calculate expectations of the form $\mathbb{E}[\sum_{i=1}^d X_i \,|\, X_k > \mathrm{VaR}_\alpha(X_k)]$, where the choices of h, g and A are trivial.

Chapter 4

Capital allocation for copula-dependent risk models

In this chapter we propose to approach the problem of sampling from a (highly) constrained multivariate distribution by (1) making explicit use of its copula and (2) sequentially approaching the target distribution through a sequence of less constrained densities. The usage of the copula structure in the sampling procedure has being independently developed in [Arbenz et al., 2014] (see also the discussion in Section 3.6.1) while the sequential approach was studied in several papers, such as [Johansen et al., 2006] and [Cérou and Guyader, 2007]. The main contribution of this chapter is to merge these two ideas and propose an algorithm which is flexible enough to solve the problem of capital allocation for generic risk distributions. Moreover we provide numerical evidence of its efficiency when compared to a simple Monte Carlo scheme or the state of the art algorithms.

Parts of this chapter were published in [Targino et al., 2015].

4.1 Reaching rare-events through sequences of intermediate sets

The idea of using intermediate sets to approximate the conditional density $f_{\mathbf{X}|g(\mathbf{X})\in A}(\mathbf{x})$ is to start sampling from the unconditional distribution $f_{\mathbf{X}}(\mathbf{x})$ and move the weighted particles towards the rare conditioning set through "not so rare" sets, as in the splitting algorithms discussed in Section 3.6

Using the notation from the previous section, for a fixed function g and set A, let $\{A_t\}_{t=1}^{T}$ be a sequence of nested sets shrinking to A, i.e., $A_t \subset A_{t-1}$ and $A_t \downarrow A$, when $t \to T$. This sequence of sets defines a sequence of regions (as before):

$$\begin{aligned}
\mathcal{G}_{\mathbf{X}_t} &:= \{\mathbf{x}_t \in \mathbb{R}^d \,:\, g(\mathbf{x}_t) \in A_t\}, \\
\mathcal{G}_{\mathbf{U}_t} &:= \left\{\mathbf{u}_t \in [0,1]^d \,:\, \left(F_1^{-1}(u_{t,1}), \ldots, F_d^{-1}(u_{t,d})\right) \in \mathcal{G}_{\mathbf{X}_t}\right\}.
\end{aligned}$$

Although it is true that

$$\mathbf{x}_t \in \mathcal{G}_{\mathbf{X}_t} \iff \mathbf{u}_t \in \mathcal{G}_{\mathbf{U}_t},$$

with $\mathbf{u}_t = (u_{t,1}, \ldots, u_{t,d}) := \big(F_1(x_{t,1}), \ldots, F_d(x_{t,d})\big)$ we will see in the sequel that working in the bounded space $[0,1]^d$ will have some advantages in the design of the algorithm. In this set-up our goal will be ultimately to have (weighted) samples $\left\{\mathbf{u}_T^{(j)}, W_T^{(j)}\right\}_{j=1}^N$ from the conditional copula $c(\mathbf{u}_T \,|\, \mathbf{u}_T \in \mathcal{G}_{\mathbf{U}_T})$ which would be then transformed through the marginal inverse c.d.f.'s in order to get a weighted sample from $f_{\mathbf{X}}(\mathbf{x} \,|\, \mathbf{x} \in \mathcal{G}_{\mathbf{X}_T})$.

Following the notation used in Section 3.4.3 we define our target distribution at each time step (level) $t = 1, \ldots, T$ as

$$\pi_t(\mathbf{u}_t) := \frac{c(\mathbf{u}_t)\mathbb{1}_{\{\mathbf{u}_t \in \mathcal{G}_{\mathbf{U}_t}\}}(\mathbf{u}_t)}{\mathbb{P}[\mathbf{U}_t \in \mathcal{G}_{\mathbf{U}_t}]}. \tag{4.1}$$

4.1.1 Copula constrained geometry

Before we formalize the algorithm to sample from the constrained copula, we study some properties of the restricted region in the copula space, defined in (3.15) and (3.16), for the particular case where $g(\mathbf{X}) = \sum_{i=1}^d X_i$ and $A = [B, +\infty)$. The idea is that the knowledge of the restricted region can help us to design more efficient sampling schemes. Similar analyses can be performed for different restrictions.

For the particular choices of $g(\cdot)$ and A given above, our interest, in $\mathbb{R}^d$, is to study points such that $\sum_{i=1}^d x_i = B$ which turn out to be equivalent to points in $[0,1]^d$ such that $\sum_{i=1}^d F_i^{-1}(u_i) = B$. It is easy to see that each of these curves (in $\mathbb{R}^d$ or in $[0,1]^d$) lie in a $d-1$ dimensional space. Formally, these curves are defined through the following mappings

$$\widetilde{\mathcal{G}}_{\mathbf{X}} := \left\{ (x_1, \ldots, x_{d-1}) \in \mathbb{R}^{d-1} \,:\, (x_1, \ldots, x_{d-1}, B - \sum_{i=1}^{d-1} x_i) \right\},$$

$$\widetilde{\mathcal{G}}_{\mathbf{U}} := \left\{ \mathbf{u}_{-d} := (u_1, \ldots, u_{d-1}) \in [0,1]^{d-1} \,:\, \big(u_1, \ldots, u_{d-1}, r(\mathbf{u}_{-d})\big) \right\}, \tag{4.2}$$

where $r(\mathbf{u}_{-d}) := F_d\Big(B - \sum_{i=1}^{d-1} F_i^{-1}(u_i)\Big)$.

First of all, note that if g is a generic continuous function and all the marginal c.d.f.'s $F_1, \ldots, F_d$ are continuous, then the curve $\widetilde{\mathcal{G}}_{\mathbf{U}}$ (defined similarly to (4.2)) will be continuous. Moreover, the region $\mathcal{G}_{\mathbf{U}}$ in (3.16) will not be the union of disjoint set, but only one continuous region. Some other properties of these regions may be derived in particular cases. For example, we know that in the linear case, i.e., $g(\mathbf{X}) = \sum_{i=1}^d X_i$, the curve in $[0,1]^d$ passes through the points $\big(F_1(B), 0, \ldots, 0\big)$, $\big(0, F_2(B), 0, \ldots, 0\big), \ldots,$ $\big(0, \ldots, 0, F_d(B)\big)$.

Another interesting information about the curve $\widetilde{\mathcal{G}}_{\mathbf{U}}$ is given by its curvature, as seen in the next Proposition.

Proposition 4.1.1. *The curve $\widetilde{\mathcal{G}}_U$ defined in (4.2) is convex at $(u_1, .., u_{d-1}, r(\boldsymbol{u}_{-d}))$ if*

$$\langle \boldsymbol{u}_{-d}, \nabla^2 r(\boldsymbol{u}_{-d}) \boldsymbol{u}_{-d} \rangle > 0, \quad \forall \boldsymbol{u}_{-d} \in \mathbb{R}^{d-1},$$

where $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ and $\nabla^2 f$ is the Hessian matrix of f.

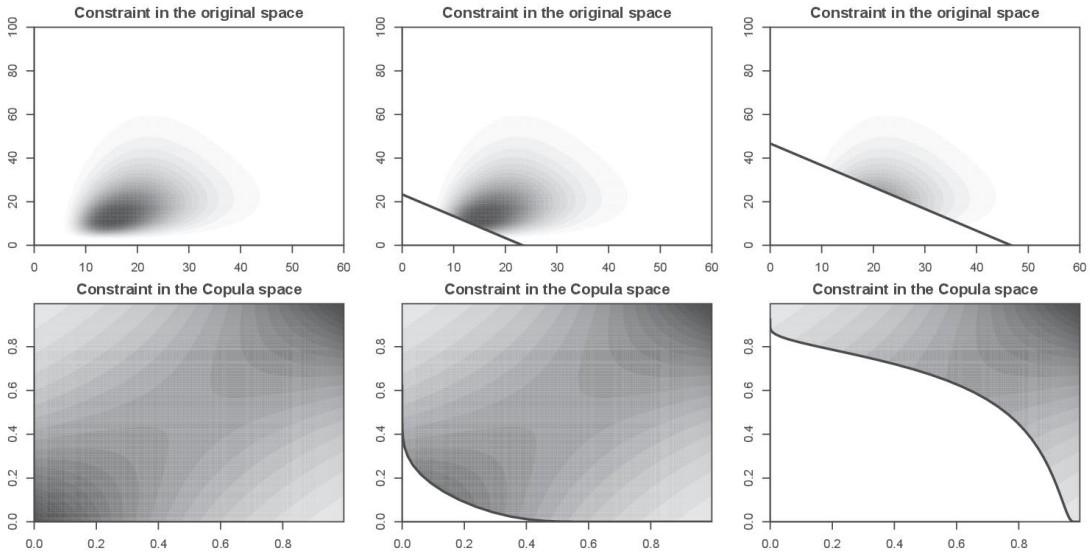

Figure 4.1: Frank copula with parameter 2, Log-Normal marginal c.d.f.'s, both with same $\mu = 3$, and $\sigma_1 = 0.4$, $\sigma_2 = 0.6$. (Top row) Constraint in the original space, for $B = 0, 23, 46$ (Bottom row) Constraint in the Copula space, $[0,1]^2$, for equivalent levels.

In the particular case where $r(\mathbf{u}_{-d}) := F_d\Big(B - \sum_{i=1}^{d-1} F_i^{-1}(u_i)\Big)$ the general terms of the Hessian matrix are given by

$$\frac{\partial^2 r}{\partial u_j \partial u_k}(\mathbf{u}_{-d}) = \frac{f_d'\left(B - \sum_{i=1}^{d-1} x_i\right)}{f_j(x_j) f_k(x_k)}, \quad \forall j \neq k, \quad j,k = 1,\dots,d-1$$

$$\frac{\partial^2 r}{\partial u_j^2}(\mathbf{u}_{-d}) = \frac{f_d'\left(B - \sum_{i=1}^{d-1} x_i\right) f_j(x_j) + f_d\left(B - \sum_{i=1}^{d-1} x_i\right) f_j'(x_j)}{\left[f_j(x_j)\right]^3},$$

$$\forall j = k, \quad j = 1,\dots,d-1$$

where, once again, we use the notation $x_i = F_i^{-1}(u_i)$ to make the above formulas more appealing.

From Proposition 4.1.1, in the very particular case where $d = 2$ and $X_1, X_2 \geq 0$ (representing losses, for example) the concavity of $\widetilde{\mathcal{G}}_{\mathbf{U}}$ is determined only by the sign of

$$f_2'\left(B - x_1\right) f_1(x_1) + f_2\left(B - x_1\right) f_1'(x_1).$$

This is due to the fact that the denominator is the power of a density function (non-negative) and that x_1 is non-negative.

On Figure 4.1 we can see that for different constraint levels the curve in $[0,1]^2$ presents different shapes, continuously varying from a convex to a concave region.

4.2 Design of a SMC sampler with linear constraints for capital allocation

In this section we return to the problem of sampling from the distribution of $(\mathbf{X} \,|\, \sum_{i=1}^{d} X_i > B)$ producing samples from $\mathbf{U} \in \mathcal{G}_{\mathbf{U}}$, as explained in section 3.5. To use the algorithm specified in Section 3.4.3 we still need to design: (1) the forward kernels $K_t(u_{t-1}, u_t)$,

(2) the backward kernels $L_{t-1}(u_t, u_{t-1})$ and (3) a Markov Chain move kernel M (in the spirit of Section 3.4.1). For the backward kernel we use the approximation to the optimal one presented in Section 3.4.3.1; the forward kernel and the "move" kernel are presented, respectively, in sections 4.2.1 and 4.2.2

4.2.1 Forward kernel

For the forward kernel $K_t(\mathbf{u}_{t-1}, \mathbf{u}_t)$, if we can guarantee that any move from $\mathbf{u}_{t-1}$ to $\mathbf{u}_t$ will already be in $\mathcal{G}_{\mathbf{U}_t}$ then we will not loose any particle in the mutation step, improving the efficiency of the algorithm. Since we are developing the sampling procedure in $[0,1]^d$ then, under the assumption that we can precisely characterize the constraint region $\mathcal{G}_{\mathbf{U}_t}$ (i.e., we can calculate F_i^{-1} for all $i = 1, \ldots, d$) then we can propose a "slice-sampling" procedure for K_t, as described below.

The idea of this type of kernel is that we can first sample $d-1$ coordinates of the vector $\mathbf{u}_t$ (chosen randomly) and then, conditional on these values, sample the last component constrained to an interval that will ensure that $\sum_{i=1}^{d} x_i > B_t$. In general these kernels will look like

$$K_t(\mathbf{u}_{t-1},\, \mathbf{u}_t) = \sum_{m=1}^{d} \left[K_t^{(-m)}(\mathbf{u}_{t-1,-m}, \mathbf{u}_{t,-m}) K_t^{(m)}(u_{t-1,m}, u_{t,m}) \right] p_m, \tag{4.3}$$

where p_m is the probability of the m-th coordinate being the last one to be chosen, $\mathbf{u}_{t,-m} = (u_{t,1}, \ldots, u_{t,m-1}, u_{t,m+1}, \ldots, u_{t,d})$ is the vector $\mathbf{u}_t$ without its m-th coordinate and $K_t^{(-m)}$ is the kernel that moves the $d-1$ dimensions of $\mathbf{u}_{t-1,-m}$ to time t. Similarly, $K_t^{(m)}$ denotes the kernel that moves $u_{t-1,m}$ to $u_{t,m}$ ensuring that $\sum_{i=1}^{d} x_{t,i} > B_t$.

To guarantee the condition is satisfied, $K_t^{(m)}$ needs to be defined over $[B_t^u(m), 1]$, where

$$B_t^u(m) := F_m^{-1}(B_t^x(m)) \tag{4.4}$$

with

$$B_t^x(m) := \max\left\{0,\, B_t - \sum_{\substack{i=1 \\ i \neq m}}^{d} F_i(u_{t,i})\right\}. \tag{4.5}$$

For simplicity, we can choose the last move to be uniformly distributed in $[B_t^u(m), 1]$, leading to

$$K_t^{(m)}(u_{t-1,m}, u_{t,m}) = \frac{u_{t,m}}{1 - B_t^u(m)} \mathbb{1}_{\{u_{t,m} \in [B_t^u(m), 1]\}}(u_{t,m}).$$

For the sake of simplicity, we only discuss the case where $K_t^{(-m)}$ consists of independent moves in each dimension, i.e.,

$$K_t^{(-m)}(\mathbf{u}_{t-1}, \mathbf{u}_t) = \prod_{\substack{i=1 \\ i \neq m}}^{d} K_t^{(-m,i)}(u_{t-1,i}, u_{t,i}). \tag{4.6}$$

Moreover, we also assume that $p_m = 1/d$, for all $m = 1, \ldots, d$, giving equal probabilities to all elements being the last one to be chosen.

Uniform moves in $\mathcal{G}_{\mathbf{U}}$ The first (naïve) idea is to define the move in each component of $\mathbf{u}$ as uniform, leading to a marginal kernels $K_t^{(-m,i)}(u_{t-1,i},\, u_{t,i}) = u_{t,i}\mathbb{1}_{\{u_{t,i}\in[0,1]\}}(u_{t,i})$ and, consequently,

$$K_t(\mathbf{u}_{t-1}, \mathbf{u}_t) = \frac{1}{d}\sum_{m=1}^{d}\left(\prod_{\substack{i=1\\ i\neq m}}^{d} u_{t,i}\mathbb{1}_{\{u_{t,i}\in[0,1]\}}(u_{t,i})\right)\left(\frac{u_{t,m}}{1-B_t^u(m)}\mathbb{1}_{\{u_{t,m}\in[B_t^u(m),1]\}}(u_{t,m})\right).$$

As we can see from the construction of this kernel, it is clearly independent of u_{t-1} and the comments by the end of Section 3.4.3.1 apply, meaning that the problem reduces to a series of Importance Sampling problems.

Global adaptive Beta moves in $[0,1]^{d-1}$ One strategy to use the information contained in $\mathbf{u}_{t-1}$ in the mutation step is to use the whole set of weighted particles at $t-1$ to estimate the parameters of the mutation kernel (subject to some restriction).

Since our kernels are defined in $[0,1]$, a reasonable idea is to use a global Beta kernel, in the sense that all particles at time $t-1$ are mutated through the same kernel. To select the parameters of the Beta distribution we match the first two moments of the Beta distributions with its sample moments at time $t-1$. Formally, let us denote $\{\mathbf{u}_{t-1}^{(j)},\, W_{t-1}^{(j)}\}_{j=1}^N$ the set of N weighted particles at time $t-1$ and $K_t^{(-m,i)}(u_{t-1,i},\, u_{t,i}) = Beta(u_{t,i};\, \alpha_{t-1,i},\, \beta_{t-1,i})$, where the right hand side term denotes the density of a random variable $Y_{t-1,i}$ which is Beta distributed with parameters $\alpha_{t-1,i}$ and $\beta_{t-1,i}$, evaluated at $u_{t,i}$. Then, matching the first two moments we have

$$\mathbb{E}[Y_{t-1,i}] = \frac{\alpha_{t-1,i}}{\alpha_{t-1,i}+\beta_{t-1,i}} = \sum_{j=1}^{N} W_{t-1}^{(j)} u_{t-1,i}^{(j)} =: \widehat{\mu}_{t-1,i},$$

$$\mathrm{Var}(Y_{t-1,i}) = \frac{\alpha_{t-1,i}\beta_{t-1,i}}{(\alpha_{t-1,i}+\beta_{t-1,i})^2(\alpha_{t-1,i}+\beta_{t-1,i}+1)} = \widehat{\mu}_{t-1,i}^2 - \sum_{j=1}^{N} W_{t-1}\left(u_{t-1,i}^{(j)}\right)^2 := \widehat{\sigma^2}_{t-1,i}$$

and after some algebra we find

$$\widehat{\alpha}_{t-1,i} = \frac{(1-\widehat{\mu}_{t-1,i})}{\widehat{\sigma^2}_{t-1,i}}\widehat{\mu}_{t-1,i}^2$$
$$\widehat{\beta}_{t-1,i} = \widehat{\alpha}_{t-1,i}\left(\frac{1}{\widehat{\mu}_{t-1,i}} - 1\right).$$

Therefore, an approximation for the mutation kernel at time t and dimension i – as in (4.6) – is given by

$$K_t^{(-m,i)}(u_{t-1,i},\, u_{t,i}) = Beta(u_{t,i};\, \widehat{\alpha}_{t-1,i},\, \widehat{\beta}_{t-1,i}). \tag{4.7}$$

Remark 4.2.1. *It is important to emphasize the kernel in (4.7), when plugged into (4.6) and (4.3), does not require any tuning and is not independent of $u_{t-1,i}$, as the parameters of the Beta distribution depend on these values.*

Figure 4.2 exemplifies the mutation of one particle $u_{t-1} = (0.2, 0.7)$ (which is in the $(t-1)$-th level set) to $u_t = (0.6, 0.9)$ (which is in the t-th level set). The mutation starts moving the first coordinate of u_{t-1} through a Beta distribution and then the second coordinate is moved following a uniform distribution defined in the appropriate region.

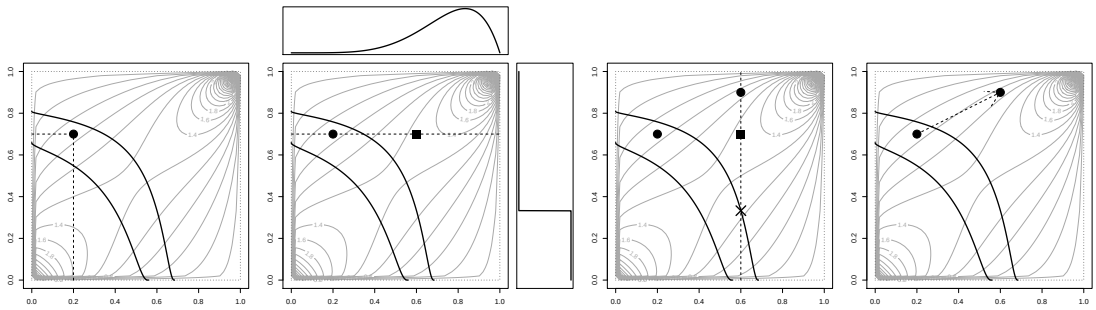

Figure 4.2: Example of the Global Beta kernel for a Gumbel(1.5) copula with Log-Normal marginals: $(\mu_1 = 0.6, \sigma_1 = 1.4)$, $(\mu_2 = 0.4, \sigma_2 = 1)$. The boundaries are such that $F_1^{-1}(u_1) + F_2^{-1}(u_2) > 2.25$ and 3.57.

4.2.2 Markov Chain move kernel

Since the forward kernels designed in Section 4.2.1 ensure the new particles will satisfy the condition at level t one possibility is to use the same kernel as a proposal in a M-H algorithm. The drawback would be that in higher dimensions the acceptance rate of the M-H would be extremely small. Instead, here we propose the use of a Gibbs sampling algorithm, that should be always preferred when the full conditional densities are known.

Suppressing the dependence in t in the vector $\mathbf{u}$ and denoting $v^*(m) := (u_1^*, \ldots, u_m^*, u_{m+1}, \ldots, u_d)$ we have that the full conditional for a generic coordinate $m = 1, \ldots, d$ can be written as

$$\begin{aligned}\pi_t(u_m^* \,|\, u_1^*, \ldots, u_{m-1}^*, u_{m+1}, \ldots, u_d) &= \frac{\pi_t(u_1^*, \ldots, u_m^*, u_{m+1}, \ldots, u_d)}{\pi_t(u_1, \ldots, u_{m-1}, u_{m+1}, \ldots, u_d)} \\ &\propto \pi_t(v^*(m)) \\ &\propto c(v^*(m)) \mathbb{1}_{\{v^*(m) \in \mathcal{G}_{\mathbf{U}_t}\}}(v^*(m)).\end{aligned}$$

Note that the full conditional distribution for the m-th coordinate of $\mathbf{u}$ is a probability distribution for u_m^*. On the other hand, since $u_1^*, \ldots, u_{m-1}^*, u_{m+1}, \ldots, u_d$ are fixed, we can rewrite the condition $[v^*(m) \in \mathcal{G}_{\mathbf{U}_t}]$ as $u_m^* \in [B^u(m), 1]$, with $B^u(m)$ as in (4.4).

To sample u_m^* from its full conditional distribution one can use a univariate slice sampler algorithm (see [Neal, 2003] and Section 3.3.3, above), which only requires the full conditional target up to a normalizing constant. In Figure 4.3 we present an example of such a Markov Chain move. On the leftmost plot, the initial point is $(u_1, u_2) = (0.9, 0.3)$. First, the support of the full conditional distribution is calculated, i.e. $B^u(1) = 0.6$, and plotted as a cross on the second figure. Then, a value $u_1^* = 0.8$ is sampled from $\pi(u_1 \,|\, u_2 = 0.3)$ (a square in the second plot). On top of the second plot we present the full conditional distribution (truncated on the left at $B^u(1) = 0.6$). For this value we find that $B^u(2) = 0$ and the support of the next full conditional distribution is $[0, 1]$ (the actual density is plotted vertically). The second coordinate $u_2^* = 0.7$ is then sampled from $\pi(u_2 \,|\, u_1^* = 0.8)$. In the last plot we have the final move, from $(u_1, u_2) = (0.9, 0.3)$ to $(u_1^*, u_2^*) = (0.8, 0.6)$.

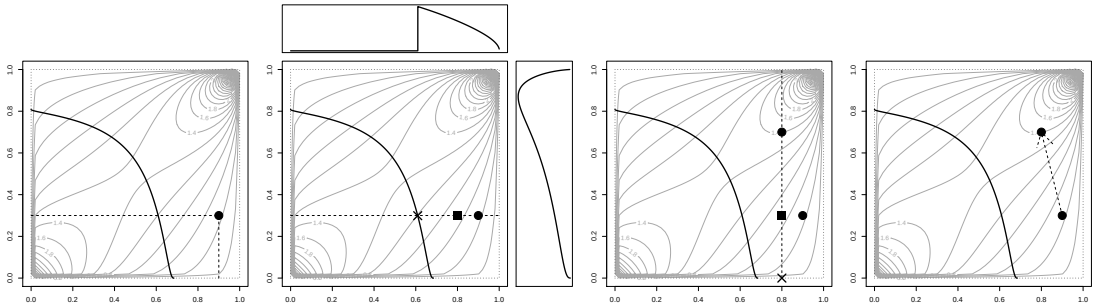

Figure 4.3: Example of the move kernel for a Gumbel(1.5) copula with Log-Normal marginals: ($\mu_1 = 0.6, \sigma_1 = 1.4$), ($\mu_2 = 0.4, \sigma_2 = 1$). The boundary is such that $F_1^{-1}(u_1) + F_2^{-1}(u_2) > 3.57$.

4.3 Case studies

In this section we present some simulation examples of the performance of the proposed Copula-Constrained SMC Sampler algorithm. For all the simulations and density calculations we made extensive use of the `R`-package `copula` [Hofert et al., 2014].

For $S = \sum_{i=1}^{d} X_i$, the aim of all methods presented here is to calculate conditional expectations of the form

$$\mathcal{A}_i = \mathbb{E}\left[X_i \,|\, S > \mathrm{VaR}_\alpha(S)\right], \quad \text{for } i = 1, \ldots, d \tag{4.8}$$

where the α quantile (VaR_α) is assumed to be perfectly known (see comments below) and $\alpha \in (0.1, 0.2, \ldots, 0.9, 0.95, 0.99, 0.995, 0.999, 0.9995, 0.9999, 0.99995)$. Moreover, as in [Arbenz et al., 2014], we assume the marginal distributions of $\mathbf{X}$ are Log-Normal with

$$X_i \sim LN(10 - 0.1i,\, 1 + 0.2i), \quad i = 1, \ldots, d.$$

For all the examples presented here, since we are not able to express the VaR_α of the aggregated process in a closed form, the first step is to to calculate a reliable approximation of this quantity, for each level α of interest. This is done through a Monte Carlo simulation of the loss vector $\mathbf{X} = (X_1, \ldots, X_d)$ – from which we can compute the aggregate loss S – of fixed size $N_q = 10,000,000$. Given a particular sample of size N_q, all the quantiles of the aggregate loss can be calculated. This process is, then, repeated for $N_{rep} = 500$ times, and the α-quantile is set as the average of the α-quantiles over all the 500 runs. The reader should note that the estimate of extreme quantiles (for example $\alpha = 0.9999$) will be less precise than the estimate of lower quantiles (such as $\alpha = 0.3$), but for the purpose of comparing the proposed algorithm with competing strategies this is irrelevant, as long as the quantile used in the conditioning argument of (4.8) is the same for all methods.

After calculating the quantiles for all levels α (which are, from now on, assumed to be exact) the baseline comparison values for the expectations in (4.8) are calculated as follows. For each level α we sample as many loss vectors $\mathbf{X}$ as necessary in order to have a sample of size $N_{MC} = 1,000$ satisfying the condition $S > \mathrm{VaR}_\alpha(S)$. At this point we note that this naive Monte Carlo sampling strategy is very inefficient and would never be utilized in practice, due

to the huge computational cost, but it provides us reference comparison to our more efficient SMC Sampler. To perform these simulations we required the usage of hundreds of cores from UCL Legion High Performance Computing Facility.

The expectations in (4.8) are, then, estimated as

$$\widehat{\mathcal{A}}_{i,MC} = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} X_i^{(j)},$$

where $\mathbf{X}^{(j)}$, $j = 1, \ldots, N_{MC}$ are the samples satisfying the condition $S > \mathrm{VaR}_\alpha(S)$. This procedure is repeated $N_{rep} = 500$ and the Monte Carlo estimate $\overline{\widehat{\mathcal{A}}}_{i,MC}$ is taken as the average over these N_{rep} repetitions, as follows

$$\overline{\widehat{\mathcal{A}}}_{i,MC} = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \widehat{\mathcal{A}}_{i,MC}^{(k)},$$

where $\widehat{\mathcal{A}}_{i,MC}^{(k)}$ stands for the estimate (using N_{MC} particles) of $\mathcal{A}_{i,MC}$ from the k-th run (out of N_{rep}). Analogously we can also define the variance of the MC estimator, $\mathrm{Var}(\widehat{\mathcal{A}}_{i,MC})$. Likewise, we denote by N_{SMC} the number of particles used in the SMC algorithm, and by $\mathrm{Var}(\widehat{\mathcal{A}}_{i,SMC})$ the variance of its estimate, also calculated using N_{rep} runs.

One may observe that the expected number of samples in the Monte Carlo scheme in order to have N_{MC} samples satisfying the α condition is equal to $M_{MC} = N_{MC}/(1-\alpha)$, which can be prohibitive if α is very close to 1.

For all the examples, the efficiency of the algorithms is measured with respect to the Variance Reduction when compared with a simple Monte Carlo scheme (properly normalized). More formally, if the SMC algorithm uses T levels to approximate (4.8) then we denote by Variance Reduction the ratio:

$$\text{Variance Reduction} = M_{MC} \times \mathrm{Var}(\widehat{\mathcal{A}}_{i,MC}) \Big/ T \times N_{SMC} \times \mathrm{Var}(\widehat{\mathcal{A}}_{i,SMC}). \tag{4.9}$$

We note this is a conservative measure of the Variance Reduction, as typically practitioners may only use in the denominator $N_{SMC} \times \mathrm{Var}(\widehat{\varphi}_{SMC})$. In addition, the Variance Reduction must be analysed in conjunction with the estimation bias. For this purpose we study the Relative Bias, defined as the relative difference from the SMC estimate to the MC estimate (assumed to be the truth, due to the very large sample sizes taken):

$$\text{Relative Bias} = \frac{\widehat{\mathcal{A}}_{i,SMC} - \widehat{\mathcal{A}}_{i,MC}}{\widehat{\mathcal{A}}_{i,MC}}$$

If the level of interest of the expectations in (4.8) is, for example, $\alpha = 0.999$ then, the SMC algorithm designed here will use as intermediate levels the quantiles $\alpha = 0.1, 0.2, \ldots, 0.9, 0.95, 0.99, 0.995$. Although expectations conditional on quantiles at lower levels, such as $0.1, \ldots, 0.9$ are not of direct interest for risk managers, as a by-product of the SMC algorithm, weighted samples from all the intermediate levels are created and all the conditional expectations can be estimated.

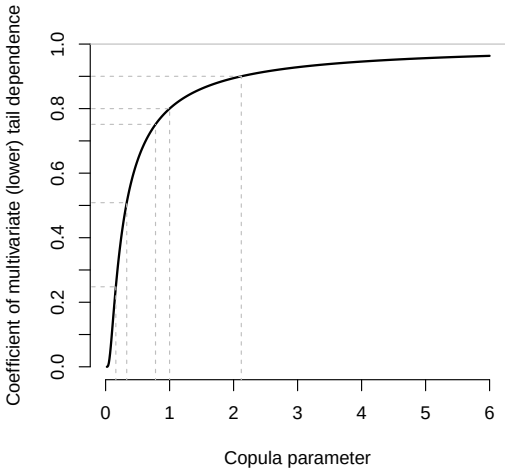

Figure 4.4: Coefficient of multivariate lower tail dependence for a 5-dimensional Clayton copula.

4.3.1 Clayton copula dependence between risk cells

In this first study we suppose we have a simple business unit and risk cell structure in which it is assumed that the dependence is on the annual losses and given simply by a Clayton copula (see Definition 2.1.8). We first study a representative simple case, with dimension $d = 5$ and investigate the behaviour of the proposed algorithm for different parameters values with fixed number of particles $N_{SMC} = 250$. To choose the parameters of interest we set the multivariate coefficient of (lower) tail dependence λ_l (see [De Luca and Rivieccio, 2012] or Proposition 2.1.20 and Figure 4.4) to be approximately equals to 0.25, 0.5, 0.75 and 0.9, which led to $\theta = 0.16$, 0.33, 0.78 and 2.12 (see Figure 4.4). To compare with the results presented in [Arbenz et al., 2014], the parameter value $\theta = 1$ is also considered.

On Figure 4.5, we present the Relative Bias (top row) and Variance Reduction (bottom row) for the entire range of different copula parameters and quantile levels. For ease of presentation, only three expectations are shown, namely the first marginal ($i = 1$), the last marginal ($i = 5$) and the sum of all marginal expectations (which is precisely the Expected Shortfall for the aggregated loss). From the Relative Bias analysis one can see that, regardless of the copula parameter and quantile levels, the estimation error is always smaller than 4% (in absolute value).

Since the estimates are unbiased, it makes sense to look at the Variance Reduction set of plots (bottom row of Figure 4.5). In the vertical axis of the plot the $\log_{10}$(Variance Reduction) is presented, meaning that, for example, the variance of the SMC algorithm is $10^{1.17} \approx 15$ times smaller than the MC scheme when $\theta = 2.12$ and $\alpha = 0.999$. The horizontal line at 0 defines the threshold where the SMC method outperforms a simple Monte Carlo: when the Variance Reduction is bellow the line the MC variance is smaller. As one should expect, for lower quantile levels a simple MC scheme should be preferred over the SMC method, since the condition in

(4.8) can be easily satisfied with a reasonably small sample size. On the other hand, as soon as the conditioning event becomes rarer, the variance of the simple Monte Carlo scheme starts to increase polynomially fast when compared with the variance of the proposed SMC algorithm.

The rarer the conditioning event the more computationally efficient it becomes to use the SMC Sampler method proposed. As previously mentioned, when $\alpha \approx 1$ the number of Monte Carlo samples required in order to generate one sample satisfying the conditioning event increases like $1/(1-\alpha)$. On the contrary, the SMC sampler is constrained to always use a fixed number of particles, independently of the rareness of the event. This is a significant advantage of such an approach.

For the ACH Importance Sampling algorithm of [Arbenz et al., 2014], discussed in Section 3.6.1, we follow the suggestion proposed in the original work involving the use of a discrete version of the optimal mixing distribution F_Λ (see (3.19)) with mass concentrated on the following 20 points: $x_k = 1-0.5^k$, $k = 1, \ldots, 20$. For a fixed quantile level α and parameter θ, the calibration of F_Λ follows the procedure proposed in [Arbenz et al., 2014, Section 6.1] which uses the stop-loss as the objective function

$$\widetilde{\Psi}(\mathbf{u}) = \max\left\{\sum_{i=1}^{d} F_i^{-1}(u_i) - \mathrm{VaR}_\alpha\left(\sum_{i=1}^{d} F_i^{-1}(u_i)\right), 0\right\}.$$

Whilst the SMC algorithm is only asymptotically unbiased (although it can be seen from Figure 4.5 the finite sample bias can be negligible) the IS–ACH is unbiased for any finite sample size $N_{IS} < \infty$. Therefore it is not necessary to analyse the Relative Bias of the method.

On the other hand, following the notation on Section 3.6.1, for a fixed parameter θ a new efficiency measure can be studied as a function of α. We denote by $P_{IS}(\alpha)$ the "percentage of particles with non-zero weight" for the α quantile. Formally this quantity is defined as

$$P_{IS}(\alpha) = \frac{\mathbb{E}[\widetilde{N}]}{\mathbb{E}[N_{\mathbf{V}}]N_{IS}}, \tag{4.10}$$

where N_{IS} is the desired sample size of the algorithm and $\widetilde{N}$ and $\mathbb{E}[N_{\mathbf{V}}]$ are, respectively, the number of particles with non-zero weight and the number of draws in the rejection algorithm in order to have one sample from $F_{\mathbf{V}}$ (see Section 3.6.1). Intuitively we should expect some Variance Reduction if, and only if, the quantity $P_{IS}(\alpha)$ is larger than $1-\alpha$.

As in the analysis made for the SMC algorithm, for the IS–ACH we also look at the (rescaled) Variance Reduction. To take into account the rejection steps in the algorithm we work with the following Variance Reduction formula

$$\text{Variance Reduction} = N_{MC} \times \widehat{\mathrm{Var}}(\widehat{\mathcal{A}}_{i,MC}) \Big/ \mathbb{E}[N_{\mathbf{V}}] \times N_{IS} \times \widehat{\mathrm{Var}}(\widehat{\mathcal{A}}_{i,IS}). \tag{4.11}$$

From Figure 4.6 (top) we can see that the percentage of particles with non-zero weight, $P_{IS}(\alpha)$, is always smaller than the $1-\alpha$, indicating an inefficiency of the IS–ACH algorithm. This inefficiency is verified in the bottom of the same figure, where the scaled Variance Reduction (as of 4.11) is presented. As in the SMC case, the Variance Reduction factor increases as a

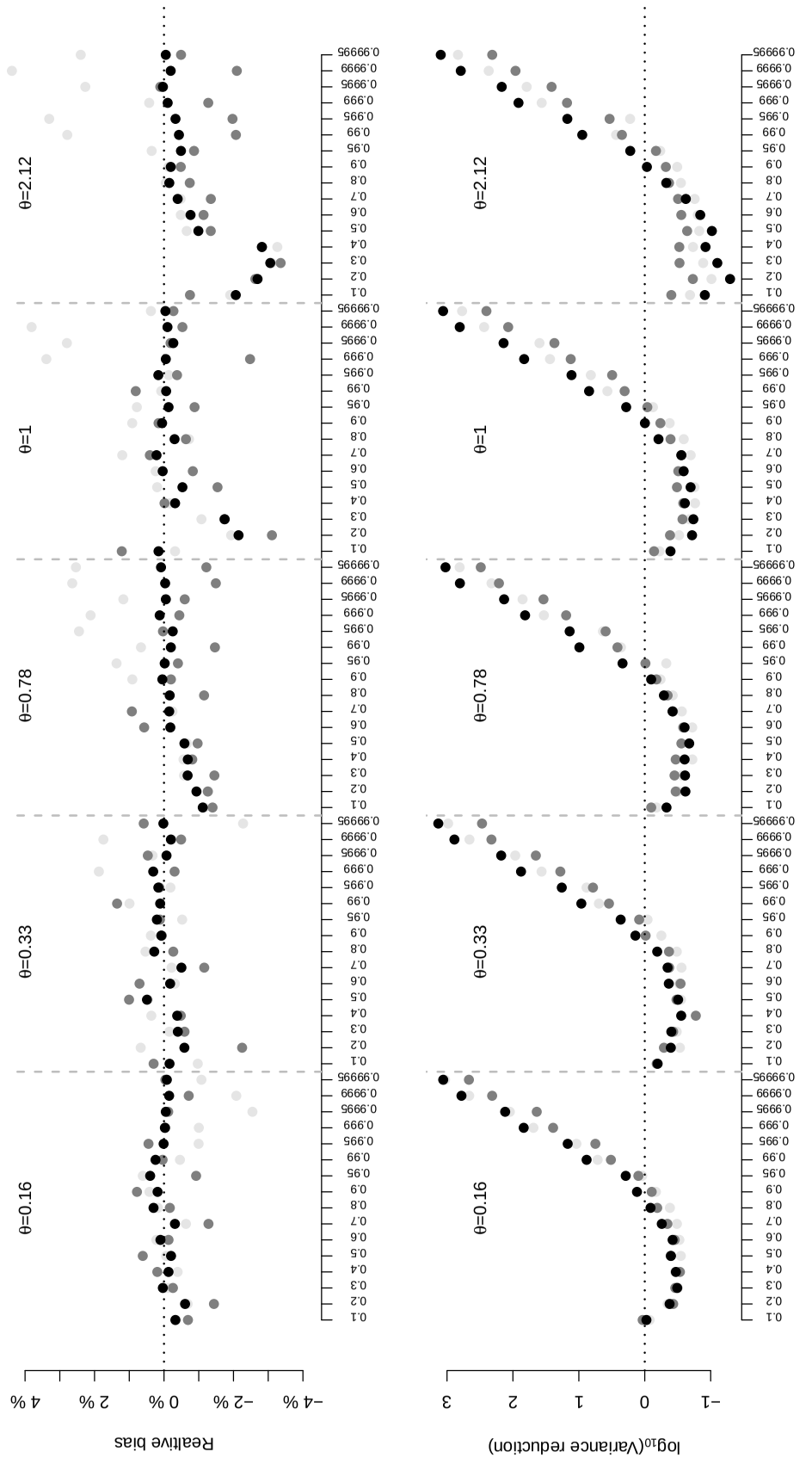

Figure 4.5: Relative Bias (top) and Variance Reduction (bottom) for the 5-dimensional Clayton copula using the SMC algorithm. Using the notation from (4.8), ● Marginal for $i = 1$, ● Marginal for $i = 5$, ● Sum of all the marginal conditional expectations (Expected Shortfall).

Figure 4.6: Ratio between the percentage of particles with non-zero weight and $1-\alpha$ (top) and Variance Reduction (bottom) for the 5-dimensional Clayton copula using the IS–ACH algorithm. Using the notation from (4.8), ● Marginal for $i=1$, ● Marginal for $i=5$, ● Sum of all the marginal conditional expectations (Expected Shortfall).

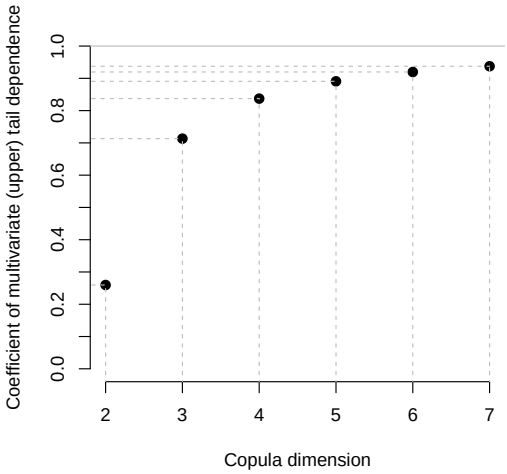

Figure 4.7: Coefficient of multivariate upper tail dependence for a Gumbel(1.25) copula.

function of α, but in the IS–ACH case it is always smaller than 1. Although this is the case, we can expect the method to be efficient (in the Variance Reduction sense) as we get even closer to $\alpha = 1$.

4.3.2 Gumbel copula dependence between risk cells

In the second example we analyse the impact of the dimension in the estimation of conditional expectations when the copula is assumed to be from the Gumbel family (see Definition 2.1.8). In this case we have chosen one parameter value ($\theta = 1.25$) in order to have values for the coefficient of multivariate upper tail dependence (see Proposition 2.1.20) ranging from very mild dependence ($\lambda_u \approx 0.25$) up to very strong dependence ($\lambda_u \approx 0.9$) in a highly constrained copula density, i.e., a single Gumbel copula in up to 7 dimensions. The coefficients of multivariate upper tail dependence are presented in Figure 4.7.

The SMC algorithm was studied for examples including dimensions $d = 2, 3, \ldots, 7$ with $N_{SMC} = 250$ particles and the results are presented on Figure 4.9. From the top row we can see the Relative Bias of the conditional expectations for low dimensional copulas (e.g., $d = 2$ or $d = 3$) is well behaved, being at most 5% of the true (Monte Carlo) value when $d = 2$ for all quantiles. When the dimensionality of the problem increases, though, a larger bias is observed. This is expected, as a single Gumbel copula in 7 dimensions, for instance, is highly constrained and its mass is mostly concentrated in a small area of the upper right quadrant of the 7-dimensional hypercube $[0, 1]^7$. In the worst case, for the first marginal of a d dimensional copula, the Relative Bias reaches more than 40% of the true value. To reduce this bias, one must increase the number of particles in the SMC Sampler. Here we have selected a very conservative set of $N_{SMC} = 250$ particles. Next we studied the bias reduction as the number of particles increases, verifying the asymptotic unbiasedness of the SMC Sampler when $N_{SMC} \to \infty$.

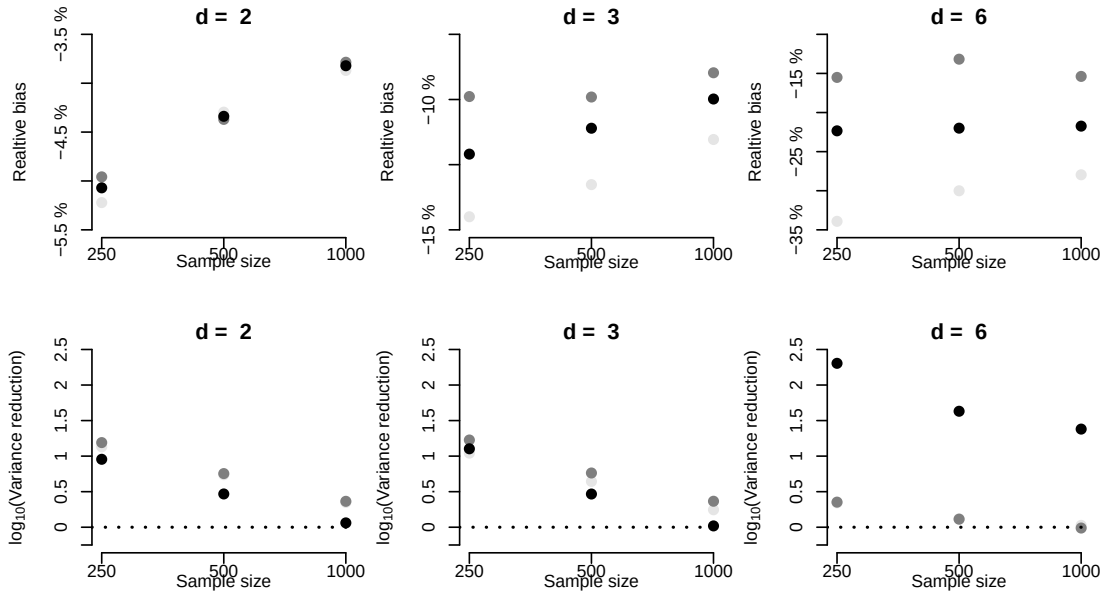

Figure 4.8: Relative Bias (top) and Variance Reduction (bottom) for the Gumbel(1.25) copula using the SMC algorithm with $N_{SMC} = 250$, 500 and $1,000$ particles. Using the notation from (4.8), $\alpha = 0.999$, • Marginal for $i = 1$, • Marginal for $i = d$, • Sum of all the marginal conditional expectations (Expected Shortfall).

From the bottom row of Figure 4.9 we can see that in all dimensions presented the SMC method is highly effective regarding decreasing the variance of the estimates when the quantile is larger than 0.999 but, as in the Clayton case, it is less efficient than a simple MC when the quantile is low.

Even though the bias involved in the SMC procedure is large for dimensions larger than $d = 3$, Figure 4.8 shows that one can decrease the absolute bias by increasing the sample size used in the SMC algorithm. For example, for the first marginal in 6 dimensions the Relative Bias goes from -35% to -30% when the number of particles increases from $N_{SMC} = 250$ to $N_{SMC} = 1,000$. The drawback of the increase in the sample size is that the method gets less effective in the Variance Reduction sense, although even in the case where $d = 6$ and $N_{SMC} = 1,000$ we still observe some humble improvement in the variance.

It is important to note that the estimation of expectations of the form (4.8) in the Gumbel model is extremely challenging, specially due to the fact that, differently from the Clayton copula, the Gumbel copula possess an intricate dependence structure near the upper right corner of the unit cube. In this case the exploration of the $[0, 1]^d$ needs to be done in an even more careful way, in order to avoid regions with low probability density. In practice, it is also important to consider the design of the mutation kernel in the SMC Sampler algorithm, if higher dimensions are considered.

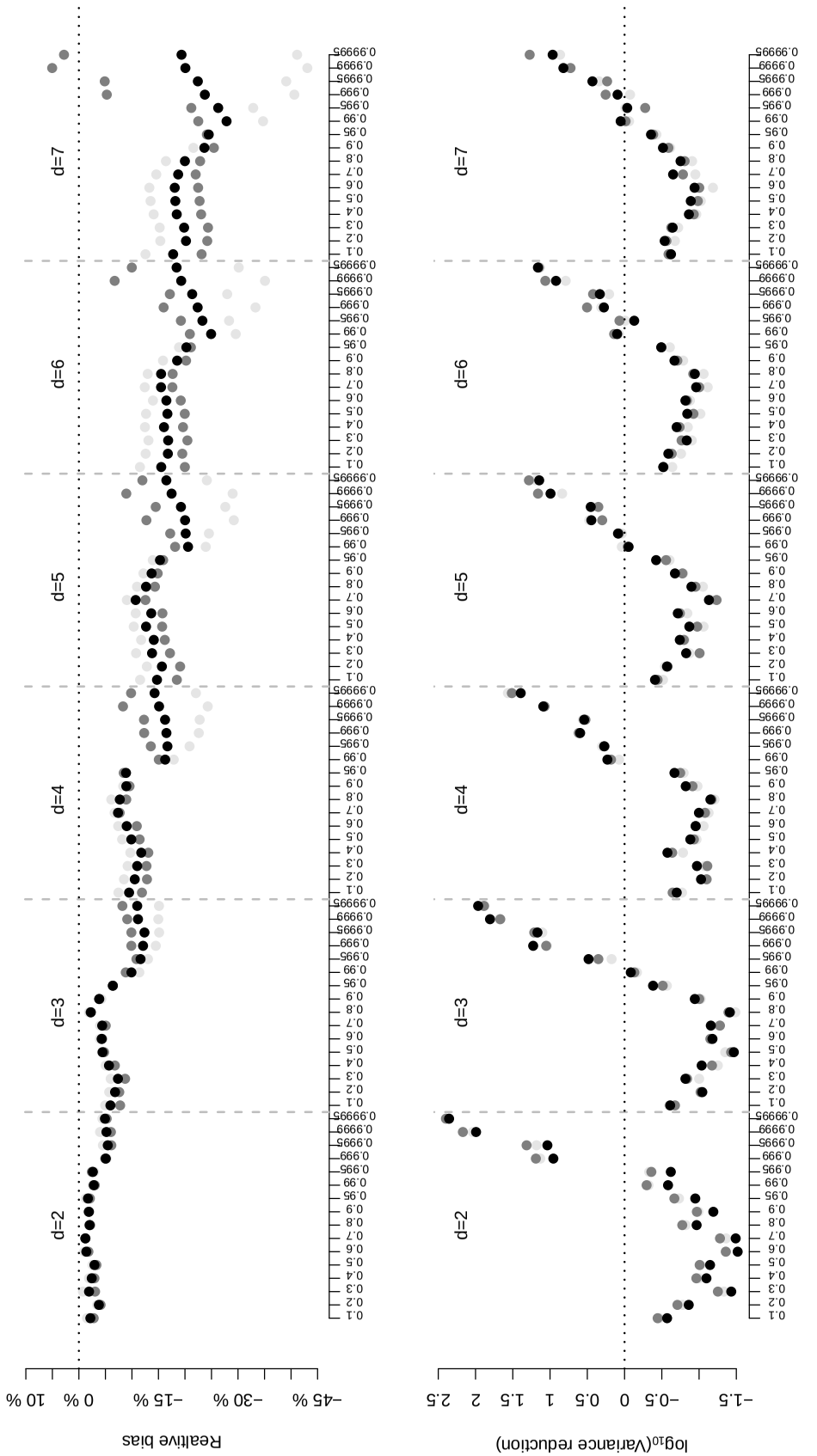

Figure 4.9: Relative Bias (top) and Variance Reduction (bottom) for the Gumbel(1.25) copula using the SMC algorithm. Using the notation from (4.8), Marginal for $i = 1$, • Marginal for $i = d$, • Sum of all the marginal conditional expectations (Expected Shortfall).

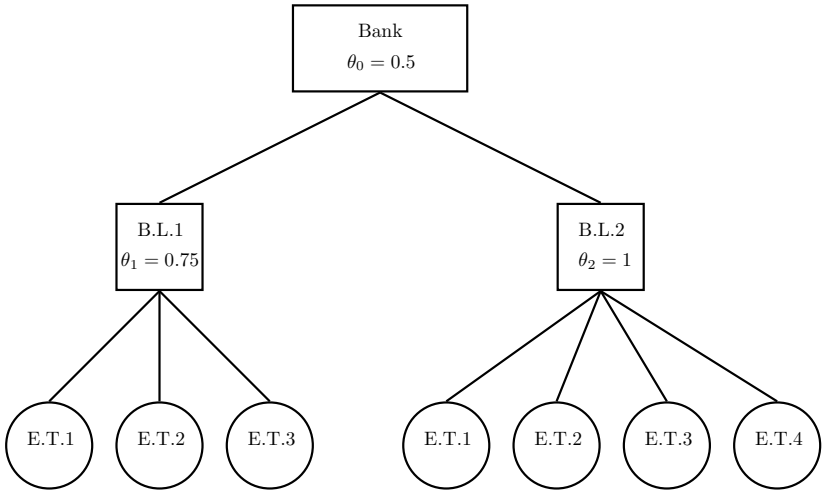

Figure 4.10: Hierarchical Clayton Copula.

4.3.3 Hierarchical Clayton copula dependence between business units and event types

As a final example we show the utilization of the SMC Sampler method in a hierarchical allocation process (as in Section 2.2.2). As a toy model, we assume a bank is divided in two different Business Units (B.U.). For the first one it is assumed that Operational Losses are due to three different Event Types (E.T.), while for B.U.2 losses may come from four different E.T.. For the simulation and also density calculation, in this example we made use the R-package `HAC` [Okhrin and Ristig, 2014].

This bank structure can be conveniently modelled with the help of a Hierarchical Archimedean Copula (HAC), also known as Nested Archimedean Copula (see [Okhrin and Ristig, 2014] and Section 2.1.1). For this example we have chosen a Hierarchical Archimedian Copula as in Figure 4.10. The dependence of the three E.T.'s on B.U.1 is given by a Clayton copula with parameter $\theta_1 = 0.75$, while within the 4 E.T.'s in B.U.2 the dependence is modelled through a Clayton copula with parameter $\theta_2 = 1$. Moreover, any loss on B.U.1 is related to losses in B.U.2 through a Clayton copula with parameter $\theta_0 = 0.5$. The copula for this model is given by

$$C(\mathbf{u}) = C_0(C_1(\mathbf{u}_1;\, \theta_1),\, C_2(\mathbf{u}_2;\, \theta_2);\, \theta_0),$$

where $C(\,\cdot\,;\theta)$ denotes a Clayton copula with parameter θ and $\mathbf{u} = (u_1, \ldots, u_7)$, $\mathbf{u}_1 = (u_1, u_2, u_3)$, $\mathbf{u}_2 = (u_4, u_5, u_6, u_7)$. The reader should note that $C_0(\,\cdot\,;\theta_0)$ is not a copula between aggregated losses. It is also important to stress the fact that this choice of parameters will ensure the Hierarchical Copula is a well defined copula, since all the members are from the same family and the parameters are decreasing from the highest to the lowest level (see, for example, [Hofert, 2010]).

As in the non-nested Clayton case, from Figure 4.11 we can see the SMC Sampler procedure

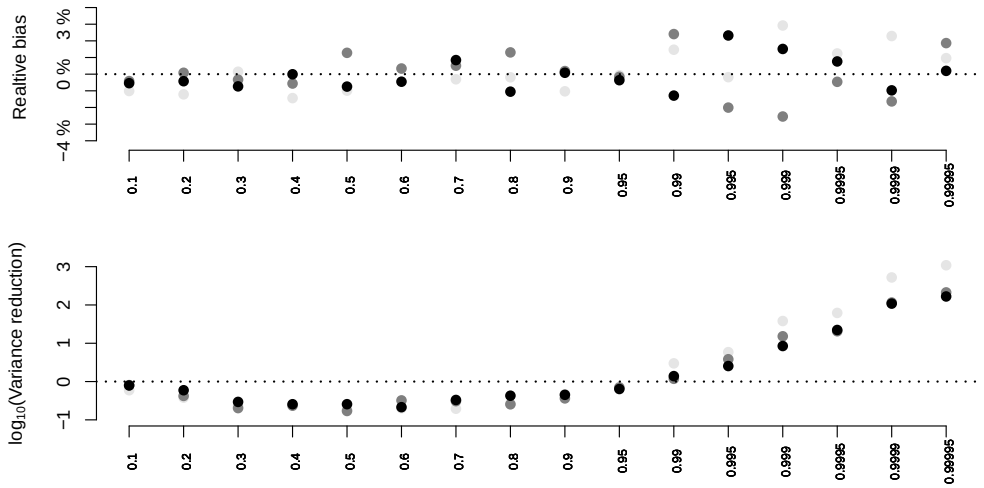

Figure 4.11: Relative Bias (top) and Variance Reduction (bottom) for the Hierarchical Clayton copula from Figure 4.10 using the SMC algorithm with $N_{SMC} = 250$ particles. Using the notation from (4.8), ● Marginal for $i = 1$, ● Marginal for $i = 7$, ● Sum of all the marginal conditional expectations (Expected Shortfall).

is unbiased for $N_{SMC} = 250$, with Relative Bias smaller than 5% in absolute terms. The method also decreases the variance of the estimates when the quantile level in the conditional event is larger than $\alpha = 0.99$, from where we can state its effectiveness.

4.4 Conclusions and final remarks

With focus on the capital allocation problem for copula-dependent risks, we presented a Sequential Monte Carlo (SMC) Sampler algorithm to calculate conditional expectations where the conditioning event is rare. We exploit the copula structure to design a SMC algorithm whose efficiency is analysed through the Variance Reduction (see Equation 4.9) when compared to a simple Monte Carlo scheme.

If, in addition to the Variance Reduction, the computational time is to be computed, one should be extremely careful, as it will be strongly dependent the programming language used, the hardware and also the actual implementation of the algorithms. For example, in `R` most of the basic functions (including sampling from simple distributions) can be used in a vectorial form, meaning that applying the function to a vector will be much faster than calculating the function values serially, i.e., inside a `for` loop. For the sake of simplicity, both in the simple MC and in the SMC schemes we sampled one value at a time and did not make use of any vectorial form.

Due to the nature of the allocation problem, there is no need for the algorithm presented to be run *online*. In most of the cases it will be used once a month or even once a year. Nevertheless, if the performance of the SMC algorithm needs to be improved, specialized libraries such as the

LibBi (see [Murray, 2013]) can be used.

Although the code used in this chapter has not been thoroughly optimised, some computational analysis can still be done. For the Clayton example from Section 4.3.1 (with parameter $\theta = 1$), for each quantile level, Table 4.4 presents (1) the time (in minutes) necessary to run the SMC algorithm; (2) how many times the SMC algorithm is slower than the simple Monte Carlo; (3) the Variance Reduction factor (as of (4.9)) for the quantities in (4.8), where ES denotes the sum of the expectations (Expected Shortfall).

			Variance Reduction		
Quantile	SMC time	$\frac{\text{SMC time}}{\text{MC time}}$	$i = 1$	$i = 5$	ES
0.1	0.32	53.83	0.34	1.95	0.46
0.2	0.64	68.97	0.27	1.16	0.29
0.3	0.97	73.94	0.11	0.14	0.37
0.4	1.30	75.04	0.07	1.89	0.19
0.5	1.63	73.09	0.19	0.16	0.18
0.6	1.97	68.15	0.35	0.03	0.28
0.7	2.31	61.41	0.42	0.21	0.78
0.8	2.65	52.55	0.30	0.38	0.21
0.9	3.00	39.28	0.44	2.94	1.83
0.95	3.35	26.34	0.29	0.33	4.10
0.99	3.70	9.64	2.79	5.36	3.54
0.995	4.04	4.56	2.98	6.54	21.87
0.999	4.39	1.28	16.56	7.20	71.37
0.9995	4.74	0.56	94.74	88.12	172.51
0.9999	5.08	0.15	224.82	154.54	272.74
0.99995	5.42	0.06	1891.81	205.48	593.32

Table 4.1: Computational time (in minutes) and Variance Reduction for the SMC algorithm when compared to a simple Monte Carlo scheme.

On this particular implementation of the algorithm, for low quantiles we can see that the MC scheme is considerably faster for the same accuracy of the SMC method. For example, when $\alpha = 0.4$ the MC method is 75 times faster than the SMC and the Variance Reduction on the first marginal expectation is 0.07, meaning that the MC variance is around 14 times smaller than the SMC variance.

On the other hand, for higher quantiles, in particular after $\alpha = 0.99$, the SMC method starts to become more appealing, since the Variance Reduction gets larger than the difference in time. It's worth noticing the proposed SMC method has been designed to be used on extreme quantiles where lower quantiles are only used as intermediate steps.

The computing times presented on Table 4.4 were measured using `R 3.1.0` in a Intel Xeon E5-1650, 3.20GHz and 16GB RAM. For each quantile level the algorithms were run 10 times and the values presented are an average over these 10 runs.

Chapter 5

A fully Bayesian risk model for the Swiss Solvency Test

As discussed in [Wüthrich, 2015], in the context of non-life insurance claims can not usually be settled immediately at the occurrence. Neither the number of claims at the current accounting year nor its amount are observable at the end of the year. The former due to claims that were incurred but not yet reported and the latter due to settlement delays. In order to be able to settle all its claims the insurance company needs to put aside sufficient provisions from the premium payments, in a process named claims reserving. The changes in these reserves over a one year period, called the claims development result (CDR), are one of the major risk drivers in the annual profit and loss statement of a non-life insurance company and, as such, play an important role in the recent solvency regulations, such as Solvency II and the Swiss Solvency Test. In order to study the impact of different lines of businesses and risk drivers in the company's solvency capital requirement (see Chapter 6) in this chapter we introduce a novel fully Bayesian model for claims reserving and discuss how to perform inference on it. On top of analyzing the one-year reserve risk, we also discuss the modelling of the one-year premium risk, as prescribed by the Swiss Solvency Test. As some of the parameters involved in these models are unknown we proposed two different approaches: a conditional and a marginalized one. In the first, most commonly encountered in the literature, we calculate quantities conditional on the unknown parameters, while under the marginalized model we integrate out the parameter uncertainty. For both models, capital allocation algorithms are proposed in Chapter 6.

Parts of this chapter are based on the working paper [Peters et al., 2016b].

5.1 The claims reserving problem

As previously mentioned, non-life insurance claims are not settled immediately at their occurrence. There is usually a *reporting delay* and also a *settlement delay* and both can vary from a couple of days to years, for more complex claims. The first type of delay may be due to administrative reasons within the company or caused (purposely or not) by the policyholder, while the latter may be due to legal investigations or the await for external information, for example.

Insurance payments start after the reporting of the claim and verification of contractual clauses and as the settlement process can take up to several decades, there is clearly an important issue with cash-flow uncertainty in this process, which is in the heart of the insurance business.

The statistical study of this uncertainty is usually performed at an aggregate level, for example, at the line of business (LoB) level, which is the approach taken in this work. Within LoBs claims are separated by their *accident year*, as claims that occur at the same year "are triggered by similar external factors, like weather conditions, economic environment, etc." [Wüthrich and Merz, 2015]. These claims, when put together, generate a cash-flow every year until their settlement. In the claims reserving literature these years are called *development years* which are usually capped at J (when we assume all the claims are settled and closed).

We assume there are L LoBs, and for each one of them we denote their (aggregated) incremental payments for claims with accident year $i = 1, \dots, I$ and development year $j = 1, \dots, J$ by $X_{i,j}^{(\ell)}$ and their cumulative payments by $C_{i,j}^{(\ell)} = \sum_{k=0}^{j} X_{i,k}$, where the $\ell = 1, \dots, L$ denotes the LoB index. Note that the incremental payments $X_{i,j}$ are made in accounting year $t = i + j$.

The information available at time $t = 0, \dots, I + J$ for the ℓ-th LoB is given by

$$\mathcal{D}^{(\ell)}(t) = \{X_{i,j}^{(\ell)} \,:\, 1 \le i \le I,\ 0 \le j \le J,\ i + j \le t\},$$

and, similarly, the total information available at time t is denoted as

$$\mathcal{D}(t) = \bigcup_{1 \le \ell \le L} \mathcal{D}^{(\ell)}(t). \tag{5.1}$$

This information is usually presented in a graphical form, called *claims triangles/trapezoids*, which may present the claims data in incremental or cumulated form. A triangle with $I = 5$ and $J = 4$ is presented in Table 5.1. The gray region (lower triangle) is called *outstanding loss liabilities*, for which the company needs to set aside some provisions, named *claims reserves*. One of key elements in the outstanding loss liabilities is the *ultimate claim*, $C_{i,J}$, which denotes the total amount to be paid for claims that occurred at year i.

Historically, in order to predict the outstanding liabilities actuaries developed deterministic algorithms, such as the chain-ladder (CL), which assumes that cumulative claims for all accident years $1 \le i \le I$ satisfy (approximately) the following relationship

$$C_{i,j+1} \approx f_j C_{i,j},$$

for (CL or *age-to-age*) factors $f_j > 0$ and $j = 0, \dots, J - 1$. To predict the ultimate claim based on the information available at time t, $\mathcal{D}(t)$, the CL algorithm prescribes a very simple rule:

$$C_{i,J} \approx C_{i,t-i} \prod_{j=t-i}^{J-1} f_j.$$

As in practice the CL factors f_j are unknown one needs to estimate them in order to compute the ultimate claim predictor. At time t, the CL algorithm prescribes the following

Accident year	Development year $j=0$	$j=1$	$j=2$	$j=3$	$j=J$
$i=1$	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,J}$
$i=2$	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	
$i=3$	$C_{3,0}$	$C_{3,1}$	$C_{3,2}$		
$i=4$	$C_{4,0}$	$C_{4,1}$			
$i=I$	$C_{I,0}$				

Table 5.1: Claims triangle/trapezoid: The upper left triangle represents the information contained in $\mathcal{D}(t)$ and the lower right triangle (in gray) represents the unknowns, i.e., $\mathcal{D}^c(t)$.

estimator for the CL factors:

$$\widehat{f}_j(t) = \frac{\sum_{k=1}^{(t-j-1)\wedge I} C_{k,j+1}}{\sum_{k=1}^{(t-j-1)\wedge I} C_{k,j}}, \tag{5.2}$$

which leads the ultimate claim $C_{i,J}$ to be predicted by

$$\widehat{\mathscr{C}}_{i,J}(t) = C_{i,t-i} \prod_{j=t-i}^{J-1} \widehat{f}_j(t). \tag{5.3}$$

In the early 1990's, by introducing randomness to the deterministic CL algorithm, the seminal paper [Mack, 1993] changed the landscape of claims reserving methods. In order to study stochastic versions of the CL algorithm we assume that $X_{i,j}$ (and consequently $C_{i,j}$) are random variables defined on the same probability space. By a slight abuse of notation we also denote by $\mathcal{D}^{(\ell)}(t)$ and $\mathcal{D}(t)$ the sigma-field generated by the corresponding sets. In this context the distribution-free model proposed in [Mack, 1993] is define as follows.

Model Assumptions 5.1.1 (Distribution-free CL model)**.** *For the distribution-free CL model we make the following assumptions.*

1. *Cumulative $C_{i,j}$ of different accident years i are independent.*
2. *There exist factors $f_0, \dots, f_{J-1} > 0$ and variance parameters $s_0^2, \dots, s_{J-1}^2$ such that for all $1 \le i \le I$ and $1 \le j \le J$ we have*

$$\mathbb{E}[C_{i,j} \,|\, C_{i,j-1}] = f_{j-1} C_{i,j-1}$$

$$\mathit{Var}(C_{i,j} \,|\, C_{i,j-1}) = s_{j-1}^2 C_{i,j-1}$$

Under Mack's CL model the ultimate claim $C_{i,J}$ is predicted by its expected value conditional on the information available at the current time, i.e., $\mathbb{E}[C_{i,J} \,|\, \mathcal{D}(t)]$ and from Model Assumptions 5.1.1 we have that

$$\mathbb{E}[C_{i,J} \,|\, \mathcal{D}(t)] = \widehat{\mathscr{C}}_{i,J}(t),$$

where the right hand sizes is defined in (5.3), the estimator based on the deterministic CL algorithm, with CL factors defined in (5.2). In [Mack, 1993] it was also proved that the estimators $\widehat{f}_j(t)$ are unbiased for f_j and uncorrelated for $0 \leq j \leq J-1$, even though $\widehat{f}_{j-1}(t)$ and $\widehat{f}_j(t)$ depend on the same data $C_{1,j}, \ldots, C_{1,(t-j)\wedge I}$.

Mack's CL model also provides a measure of uncertainty of the ultimate claim, based on its variance, but before presenting this measure we first discuss the estimator for the variance parameters s_j^2 proposed in [Mack, 1993]. Based on the information available at time $t = I$ Mack derived the following unbiased estimator for s_j^2:

$$\widehat{s}_j^2(t) = \frac{1}{t-j-2} \sum_{k=1}^{t-j-1} C_{k,j} \left(\frac{C_{k,j+1}}{C_{k,j}} - \widehat{f}_j(t) \right)^2, \ \forall 0 \leq j \leq (J-1) \wedge (I-3).$$

For $j = J-1 = I-2$ due to the lack of data in the triangle [Mack, 1993] empirically observed the exponential decay of $\widehat{s}_0^2(t), \ldots, \widehat{s}_{J-2}^2(t)$ and proposed

$$\widehat{s}_{J-1}^2(t) = \min\{\widehat{s}_{J-3}^2(t),\, \widehat{s}_{J-2}^2(t),\, \widehat{s}_{J-2}^4(t)/\widehat{s}_{J-3}^2(t)\} = \min\{\widehat{s}_{J-3}^2(t),\, \widehat{s}_{J-2}^4(t)/\widehat{s}_{J-3}^2(t)\}. \quad (5.4)$$

The uncertainty in the ultimate claim was, then, calculated as

$$\mathrm{Var}(C_{i,J} \,|\, \mathcal{D}(t)) = \left(\widehat{\mathscr{C}}_{i,J}(t) \right)^2 \sum_{j=t-i}^{J-1} \frac{\widehat{s}_j^2 / \widehat{f}_j(t)^2}{\widehat{\mathscr{C}}_{i,j}(t)}. \quad (5.5)$$

The variance formula described in (5.5) presents a *long term* view of the uncertainty involved on claims reserving and does not address the *short term* view of modern solvency regulations, such as the Swiss Solvency Test and Solvency II. Under these new regulatory frameworks *solvency* is analyzed in a one year horizon and to this end different quantities need to be studied. These quantities are described in the next sections under what we call the *marginalized* and the *conditional* models, which are distinct ways to treat the parameter uncertainty involved in the claims' models.

5.2 Claims reserve and the Swiss solvency test

In this section we present the cornerstone of the models discussed in the remaining of Chapter 5 and all of Chapter 6: the approaches we named *marginalized* and *conditional* models for claims data.

Although we postpone the construction of the specific claims payments model to Section 5.3 (where we introduce the Gamma-Gamma model) we now assume its behaviour is given by a Bayesian model depending on two parameter vectors, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, for which prior distributions are assigned. Probabilistic statements, such as the calculation of the risks allocated to each trigger, have to be made based only on the available data, denoted as $\mathcal{D}(t)$ and formally defined in Section 5.2. This requirement implies that the uncertainty on the parameter values needs to be integrated out and, for the Bayesian model postulated in this work, that can be partially done analytically. Under our model assumptions, the parameter vector $\boldsymbol{\phi}$ can be handled analytically, but the same is not possible for $\boldsymbol{\theta}$.

Therefore, to calculate the risk allocations we approximate the stochastic behaviour of functions of future observations, with the functions defined in Section 5.2. For the moment, let us denote by $\overline{\boldsymbol{Z}}$ a multivariate function of $\mathcal{D}(t+1)$ and $\boldsymbol{\theta}$ denote, respectively, the future data and the vector of model parameters. On the one hand, in the *conditional model*, we approximate the distribution of the components of the vector $\overline{\boldsymbol{Z}} \,|\, \boldsymbol{\theta}$, $\mathcal{D}(t)$. On the other hand, in the *marginalized model*, the approximation is performed after the parameter uncertainty has been integrated out (i.e., marginalized). In this later framework, we approximate the distribution of the components of $\boldsymbol{Z} \,|\, \mathcal{D}(t)$, where the random vector $\boldsymbol{Z}$ is defined as $\boldsymbol{Z} = \mathbb{E}[\overline{\boldsymbol{Z}} \,|\, \mathcal{D}(t)]$, with expectation taken with respect to $\boldsymbol{\theta} \,|\, \mathcal{D}(t)$. Note that, given $\mathcal{D}(t)$, $\boldsymbol{Z}$ is a random variable, as it depends on future information, i.e., $\mathcal{D}(t+1)$. Both in the conditional and in the marginalized models we use moment matching (with the original moments calculated from the Gamma-Gamma model of Section 5.3) and log-normal distributions for the approximations and couple the distributions via a Gaussian copula.

Suppressing the dependence on the available information, $\mathcal{D}(t)$, these two models (marginalized and conditional) are defined through their probability density functions (p.d.f.'s), $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $f_{\overline{\boldsymbol{Z}}|\boldsymbol{\theta}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta})$, respectively, which are both assumed to be combinations of log-normal distributions and a gaussian copula. For the conditional model, as we work in a Bayesian framework, the unknown parameter vector $\boldsymbol{\theta}$ has a (posterior) distribution with p.d.f. $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. This is, then, combined with the likelihood $f_{\overline{\boldsymbol{Z}}|\boldsymbol{\theta}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta})$ to construct $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$, the density used for inference under the conditional model.

For the methodology discussed in this work, the important features of these two models are that $f_{\boldsymbol{Z}}(\boldsymbol{z})$ is known in closed form, whilst $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ is not.

In summary, the two models presented in Sections 5.2 to 5.5 are defined as

$$\text{Marginalized model: } \boldsymbol{Z} \sim f_{\boldsymbol{Z}}(\boldsymbol{z}); \tag{5.6}$$

$$\text{Conditional model: } \overline{\boldsymbol{Z}} \sim f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}) = \int f_{\overline{\boldsymbol{Z}} \,|\, \boldsymbol{\theta}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{5.7}$$

Remark 5.2.1. *As the "original" model for claims payments (i.e., the Gamma-Gamma model of Section 5.3) is a Bayesian model, we use the Bayesian nomenclature for both the marginalized and the conditional model. For the former, the Bayesian structure of prior and likelihood is hidden in equation (5.6), as the parameter $\boldsymbol{\theta}$ has already been marginalized (with respect to its posterior distribution). For the later, we explicitly make use of the posterior distribution of $\boldsymbol{\theta}$ in (5.7). Another strategy, followed in [Wüthrich, 2015], is to use an "empirical Bayes" approach, fixing the value of the unknown parameter vector $\boldsymbol{\theta}$, for example at its maximum likelihood estimator (MLE).*

Although the original model for claims payments is only presented in Sections 5.3—5.5 as the quantities of interest for the capital allocation problem are independent of the model itself, they are introduced in current section.

5.2.1 Conditional predictive model

As previously mentioned, the models for claims payments in a particular LoB ℓ will involve a set of parameters, denoted as $\boldsymbol{\theta}^{(\ell)}$. For the ease of exposition, whenever a quantity is defined conditional on $\boldsymbol{\theta}^{(\ell)}$ it is going to be denoted with a bar on top of it.

At time $t \geq I$, LoB ℓ and accident year $i > t - J$ predictors for the ultimate claim $C_{i,J}^{(\ell)}$ and the corresponding claims reserves are defined, respectively as

$$\overline{\mathscr{C}}_{i,J}^{(\ell)}(t) = \mathbb{E}[C_{i,J}^{(\ell)} \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)] \quad \text{and} \quad \overline{\mathcal{R}}_{i}^{(\ell)}(t) = \overline{\mathscr{C}}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)} \tag{5.8}$$

Under the modern solvency regulations, such as Solvency II and the Swiss Solvency Test an important variable to be analysed is the claims development result (CDR), which is the difference in the ultimate claim prediction at time t and $t+1$. For accident year $i = 1, \ldots, I$, accounting year $t + 1 > I$ and LoB ℓ, the CDR is defined as

$$\begin{aligned} \overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) &= \overline{\mathcal{R}}_{i}^{(\ell)}(t) - \big(X_{i,t-i+1}^{(\ell)} + \overline{\mathcal{R}}_{i}^{(\ell)}(t+1)\big) \\ &= \overline{\mathscr{C}}_{i,J}^{(\ell)}(t) - \overline{\mathscr{C}}_{i,J}^{(\ell)}(t+1) \end{aligned} \tag{5.9}$$

and an application of the tower property of the expectation shows that

$$\mathbb{E}[\overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)] = 0. \tag{5.10}$$

Remark 5.2.2. *It should be noted that the predictor defined in (5.8) might not be optimal, since it only uses information related to the ℓ-th LoB, $\mathcal{D}^{(\ell)}(t)$, instead of all the information available, $\mathcal{D}(t)$. We proceed with the definition in (5.8) in order to avoid intractability of the models to be considered.*

Equation (5.10) justifies the prediction of the CDR by zero and the uncertainty of this prediction can be assessed by the conditional mean squared error of prediction (msep):

$$\begin{aligned} \mathrm{msep}_{\overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)}(0) &= \mathbb{E}[(\overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) - 0)^2 \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)] \qquad (5.11) \\ &= \mathrm{Var}\big(\overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)\big) \\ &= \mathrm{Var}\big(\overline{\mathscr{C}}_{i,J}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)\big). \qquad (5.12) \end{aligned}$$

Moreover, we denote the aggregated (over accident years) CDR and the reserves, conditional on the knowledge of the parameter $\boldsymbol{\theta}^{(\ell)}$, respectively, by

$$\overline{\mathrm{CDR}}^{(\ell)}(t+1) = \sum_{i=1}^{I} \overline{\mathrm{CDR}}_{i}^{(\ell)}(t+1) \quad \text{and} \quad \overline{\mathcal{R}}^{(\ell)}(t) = \sum_{i=1}^{I} \overline{\mathcal{R}}_{i}^{(\ell)}(t). \tag{5.13}$$

Using this notation we also define the total prediction uncertainty incurred when predicting $\overline{\mathrm{CDR}}^{(\ell)}(t+1)$ by zero as

$$\mathrm{msep}_{\overline{\mathrm{CDR}}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)}(0) = \mathrm{Var}\Big(\sum_{i=1}^{I} \overline{\mathscr{C}}_{i,J}^{(\ell)}(t+1) \,|\, \boldsymbol{\theta}^{(\ell)}, \mathcal{D}^{(\ell)}(t)\Big).$$

Remark 5.2.3. *It should be noticed that, in general, as the parameter vector $\boldsymbol{\theta}^{(\ell)}$ is unknown none of the quantities presented in this section can be directly calculated unless an estimate for the parameter is used.*

5.2.2 Marginalized predictive model

Even though the original cumulative claims model is defined conditional on unobserved parameter values, any quantity calculated based on this model should only depend on observable quantities. As we follow the Bayesian paradigm, unknown quantities are modelled using a priori probability distribution, which reflects prior beliefs about these parameters.

Analogously to Section 5.2.1 we define the marginalized (Bayesian) ultimate claim predictor and its reserves, respectively, as

$$\mathscr{C}_{i,J}^{(\ell)}(t) = \mathbb{E}[C_{i,J}^{(\ell)} \,|\, \mathcal{D}^{(\ell)}(t)] = \mathbb{E}_{\boldsymbol{\theta}^{(\ell)}}[\overline{\mathscr{C}}_{i,J}^{(\ell)}(t) \,|\, \mathcal{D}^{(\ell)}(t)] \quad \text{and} \quad \mathcal{R}_i^{(\ell)}(t) = \mathscr{C}_{i,J}^{(\ell)}(t) - C_{i,t-i}^{(\ell)}. \tag{5.14}$$

We also define the marginalized CDR and notice, again using the tower property, that its mean is equal to zero

$$\mathrm{CDR}_i^{(\ell)}(t+1) = \mathscr{C}_{i,J}^{(\ell)}(t) - \mathscr{C}_{i,J}^{(\ell)}(t+1) \quad \text{with} \quad \mathbb{E}[\mathrm{CDR}_i^{(\ell)}(t+1) \,|\, \mathcal{D}^{(\ell)}(t)] = 0.$$

Furthermore, summing over all accident years i we follow (5.13) and denote by $\mathcal{R}^{(\ell)}(t)$ and $\mathrm{CDR}^{(\ell)}(t+1)$ the aggregated version of the marginalized reserves and CDR, where the uncertainty in the latter is measured via

$$\mathrm{msep}_{\mathrm{CDR}^{(\ell)}(t+1) \,|\, \mathcal{D}^{(\ell)}(t)}(0) = \mathrm{Var}\Big(\sum_{i=1}^{I} \mathscr{C}_{i,J}^{(\ell)}(t+1) \,|\, \mathcal{D}^{(\ell)}(t) \Big). \tag{5.15}$$

5.2.3 Solvency capital requirement (SCR)

In this section we discuss how two important concepts in actuarial risk management, namely the technical result (TR) and the solvency capital requirement (SCR), can be defined for both the conditional and the marginalized models.

In this context, the TR is calculated netting all income and expenses arising from the LoBs, while the SCR denotes the minimum capital required by the regulatory authorities in order to cover the company's business risks. More precisely, the SCR for year $t+1$ quantifies the risk of having a substantially distressed financial result at time $t+1$, evaluated in light of the available information at time t.

As an important shorthand notation, we introduce three sets of random variables, representing the total claim amounts of the current year (CY) and of prior year (PY), the latter for both the conditional and marginalized models. These random variables are defined, respectively, as

$$Z_{CY}^{(\ell)} = \mathscr{C}_{t+1,J}^{(\ell)}(t+1), \;\; \overline{Z}_{PY}^{(\ell)} = \sum_{i=1}^{I} \Big(\overline{\mathscr{C}}_{i,J}^{(\ell)}(t+1) - C_{i,t-i}^{(\ell)} \Big) \;\text{ and }\; Z_{PY}^{(\ell)} = \sum_{i=1}^{I} \Big(\mathscr{C}_{i,J}^{(\ell)}(t+1) - C_{i,t-i}^{(\ell)} \Big). \tag{5.16}$$

In the standard SST model, CY claims do not depend on any unknown parameters and are split into L small claims, $Z_{CY,s}^{(\ell)}$ (also called attritional claims), and P claims caused due to large claim events, $Z_{CY,l}^{(p)}$. In this context the company can choose thresholds $\beta^{(\ell)}$ such that claims larger than these amounts are classified as large claims in its respective LoBs.

To further simplify the notation we group all the random variables related to the conditional and the marginalized models in two random vectors, defined as follows

$$\overline{\boldsymbol{Z}} = (\overline{Z}_1, \ldots, \overline{Z}_{2L+P}) = (\overline{Z}^{(1)}_{PY}, \ldots, \overline{Z}^{(L)}_{PY}, Z^{(1)}_{CY,s}, \ldots, Z^{(L)}_{CY,s}, Z^{(1)}_{CY,l}, \ldots, Z^{(P)}_{CY,l}), \tag{5.17}$$

$$\boldsymbol{Z} = (Z_1, \ldots, Z_{2L+P}) = (Z^{(1)}_{PY}, \ldots, Z^{(L)}_{PY}, Z^{(1)}_{CY,s}, \ldots, Z^{(L)}_{CY,s}, Z^{(1)}_{CY,l}, \ldots, Z^{(P)}_{CY,l}). \tag{5.18}$$

Next we give more details on how the TR and the SCR are calculated in the generic structure of the conditional and the marginalized models.

5.2.3.1 SCR for the conditional model

At time $t+1$ the technical result (TR) of the ℓ-th LoB in accounting year $(t, t+1]$ based on the conditional model is defined as the following $\mathcal{D}^{(\ell)}(t+1)$–measurable random variable:

$$\overline{\mathrm{TR}}^{(\ell)}(t+1) = \Pi^{(\ell)}(t+1) - K^{(\ell)}(t+1) - \mathscr{C}^{(\ell)}_{t+1,J}(t+1) + \overline{\mathrm{CDR}}^{(\ell)}(t+1),$$

where $\Pi^{(\ell)}(t+1)$ and $K^{(\ell)}(t+1)$ are, respectively, the earned premium and the administrative costs of accounting year $(t, t+1]$. For simplicity, we assume that these two quantities are known at time t. Having them random adds a small additional complexity that can also be packed into the randomness of $\mathscr{C}^{(\ell)}_{t+1,J}(t+1)$. It should also be noticed that in this context $\mathcal{D}(t)$ not only includes the information defined in (5.1), but it should rather be replaced by $\mathcal{F}(t)$, a sigma-field generated by the inclusion of the information about $\Pi(t+1)$ and $K(t+1)$, as the premium and administrative costs of accounting year $(t, t+1]$ should be predictable and, hence, $\mathcal{F}(t)$-measurable.

Given the technical result for all the LoBs, the company's overall TR based on the conditional model, aggregated cost and premium are denoted, respectively, by

$$\overline{\mathrm{TR}}(t+1) = \sum_{\ell=1}^{L} \overline{\mathrm{TR}}^{(\ell)}(t+1), \quad \Pi(t+1) = \sum_{\ell=1}^{L} \Pi^{(\ell)}(t+1) \quad \text{and} \quad K(t+1) = \sum_{\ell=1}^{L} K^{(\ell)}(t+1).$$

In order to cover the company's risks over an horizon of one year, the Swiss Solvency Test is concerned with the 99% expected shortfall (in light of all the data up to time t):

$$\overline{\mathrm{SCR}}(t+1) = \mathrm{ES}_{99\%}[-\overline{\mathrm{TR}}(t+1) \mid \mathcal{D}(t)],$$

where $\overline{\mathrm{SCR}}$ denotes the solvency capital requirement.

It is very important to notice that even though the expected shortfall operator is being applied to a "conditional random variable", namely $\overline{\mathrm{TR}}$, the operator is *not* being taken conditional on the knowledge of $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(L)})$, otherwise this quantity would not be computable (as discussed in Remark 5.2.3). Instead, the SCR is calculated based on the marginalized version of the conditional model, where the parameter uncertainty is integrated out. More precisely, the expected shortfall is based on the following (usually intractable) distribution

$$f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \mid \mathcal{D}(t)) = \int f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \mid \boldsymbol{\theta}, \mathcal{D}(t)) \pi(\boldsymbol{\theta} \mid \mathcal{D}(t)) d\boldsymbol{\theta}.$$

In order to compute the SCR based on the conditional model we first discuss the measurablity of the terms in the conditional TR, which can be rewritten as

$$\overline{\mathrm{TR}}(t+1) = K(t+1) - \Pi(t+1) - \sum_{\ell=1}^{L}\sum_{i=1}^{I}\left(C_{i,t-i}^{(\ell)} - \overline{\mathscr{C}}_{i,J}^{(\ell)}(t)\right) - \sum_{\ell=1}^{L}\left(\overline{Z}_{PY}^{(\ell)} + \overline{Z}_{CY}^{(\ell)}\right).$$

From the above equation we see the first two terms are, by assumption, $\mathcal{D}(t)$ measurable and so are all the terms of the form $C_{i,t-i}^{(\ell)}$, while the last summation is $\mathcal{D}(t+1)$ measurable and, therefore, a random variable at time t. Due to the dependence on the unknown parameter $\boldsymbol{\theta}$, the conditional ultimate claim predictor $\overline{\mathscr{C}}_{i,J}^{(\ell)}(t)$ is usually *not* $\mathcal{D}(t)$ measurable. However, under the models introduced in Section 5.3 we have that $\overline{\mathscr{C}}_{i,J}^{(\ell)}(t)$ depends only on the claims data up to time t and not on the unknown parameter vector, making it $\mathcal{D}(t)$ measurable. In this case

$$\overline{\mathrm{SCR}}(t+1) = K(t+1) - \Pi(t+1) - \sum_{\ell=1}^{L}\overline{\mathcal{R}}^{(\ell)}(t) + \mathrm{ES}_{99\%}\left[\sum_{\ell=1}^{L}\overline{Z}_{PY}^{(\ell)} + \overline{Z}_{CY}^{(\ell)} \;\middle|\; \mathcal{D}(t)\right]. \tag{5.19}$$

5.2.3.2 SCR for the marginalized model

As the parameter uncertainty is dealt with in a previous step, the calculation of the SCR for the marginalized model is simpler than its conditional counterpart.

Similarly to the conditional case, we define the TR for the marginalized model as

$$\mathrm{TR}^{(\ell)}(t+1) = \Pi^{(\ell)}(t+1) - K^{(\ell)}(t+1) - \mathscr{C}_{t+1,J}^{(\ell)}(t+1) + \mathrm{CDR}^{(\ell)}(t+1),$$

and its aggregated version as

$$\mathrm{TR}(t+1) = \sum_{\ell=1}^{L}\mathrm{TR}^{(\ell)}(t+1).$$

Furthermore, the SCR for the marginalized model is given by

$$\mathrm{SCR}(t+1) = \mathrm{ES}_{99\%}[-\mathrm{TR}(t+1) \mid \mathcal{D}(t)] \tag{5.20}$$

$$= K(t+1) - \Pi(t+1) - \sum_{\ell=1}^{L}\mathcal{R}^{(\ell)}(t) + \mathrm{ES}_{99\%}\left[\sum_{\ell=1}^{L}Z_{PY}^{(\ell)} + Z_{CY}^{(\ell)} \;\middle|\; \mathcal{D}(t)\right], \tag{5.21}$$

where in this case the expected shortfall is calculated with respect to the density $f_{\boldsymbol{Z}}(\boldsymbol{z})$.

Remark 5.2.4. *As we assume the cost of claims processing and assessment $K(t+1)$ and premium $\Pi(t+1)$ are known at time t they do not differ from the conditional to the marginalized model.*

5.3 Modelling of individual LoBs PY claims

For the modelling of the PY claims risk we need to model $\overline{Z}_{PY}$ or Z_{PY} as given in (5.16). The uncertainty in these random variables will be assessed by the conditional and marginalized mean square error of prediction (msep), introduced in (5.12) and (5.15). In order to calculate the msep we must first expand our analysis to the study of the claims reserving uncertainty. To do so, in this section we present a fully Bayesian version of the gamma-gamma chain-ladder model, which has been studied in [Peters et al., 2017].

Since in this section we present the model for individual LoBs, for notational simplicity we omit the index (ℓ) from all random variables and parameters.

Model Assumptions 5.3.1 (Gamma-Gamma Bayesian CL model)**.** *We make the following assumptions:*

(a) Conditionally, given $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{J-1})$ *and* $\boldsymbol{\sigma} = (\sigma_0, \ldots, \sigma_{J-1})$*, cumulative claims* $(C_{i,j})_{j=0,\ldots,J}$ *are independent (in accident year* i*) Markov processes (in development year* j*) with*

$$C_{i,j+1} \mid C_{i,j},\ \phi_j,\ \sigma_j \sim \Gamma\big(C_{i,j}\sigma_j^{-2},\ \phi_j\sigma_j^{-2}\big),$$

for all $1 \le i \le I$ *and* $0 \le j \le J-1$*.*

(b) The parameter vectors $\boldsymbol{\phi}$ *and* $\boldsymbol{\sigma}$ *are, a priori, independent.*

(c) For given hyper-parameters $f_j > 0$ *the components of* $\boldsymbol{\phi}$ *are independent such that*

$$\phi_j \sim \lim_{\gamma_j \to 1} \Gamma\big(\gamma_j,\ f_j(\gamma_j - 1)\big),$$

for $0 \le j \le J-1$*, where this limits infers that they are eventually distributed from an improper uninfomative prior. This assumes we are taking a objective Bayesian view.*

(d) The components σ_j *of* $\boldsymbol{\sigma}$ *are independent and* F_{σ_j}*-distributed, having support in* $(0, d_j)$ *for given constants* $0 < d_j < \infty$ *for all* $0 \le j \le J-1$*.*

(e) $\boldsymbol{\phi}$*,* $\boldsymbol{\sigma}$ *and* $C_{1,0}, \ldots, C_{I,0}$ *are independent and* $\mathbb{P}[C_{i,0} > 0] = 1$*.*

From Model Assumptions 5.3.1 (a), conditional on a specific value of the parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$, we have that

$$\begin{aligned} \mathbb{E}[C_{i,j+1} \mid C_{i,j},\ \phi_j,\ \sigma_j] &= \phi_j^{-1} C_{i,j}, \\ \mathrm{Var}(C_{i,j+1} \mid C_{i,j},\ \phi_j,\ \sigma_j) &= \phi_j^{-2}\sigma_j^2 C_{i,j}, \end{aligned} \tag{5.22}$$

which provides a stochastic, fully Bayesian, formulation of the classical CL model of [Mack, 1993].

In Model Assumptions 5.3.1 (c) the (improper) prior distribution for $\boldsymbol{\phi}$ should be seen as a non-informative limit when $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{J-1}) \to \mathbf{1} = (1, \ldots, 1)$ of the (proper) prior assumption

$$\phi_j \sim \Gamma\big(\gamma_j,\ f_j(\gamma_j - 1)\big).$$

Even though the prior is assumed improper and does not integrate to one, the conditional posterior for $\phi_j \mid \sigma_j, \mathcal{D}(t)$ is proper and, in addition, also gamma distributed, as seen in the theorem below (see also [Merz and Wüthrich, 2015, Lemma 3.2]).

Theorem 5.3.2. *Under Model Assumptions 5.3.1 the conditional posterior for* $\phi_j \mid \sigma_j, \mathcal{D}(t)$ *is given by*

$$\phi_j \mid \boldsymbol{\sigma}, \mathcal{D}(t) \sim \Gamma(a_j, b_j) \tag{5.23}$$

with the following parameters

$$a_j = 1 + \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j}\sigma_j^{-2} \quad \textit{and} \quad b_j = \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}\sigma_j^{-2}. \tag{5.24}$$

Proof. Under Model Assumptions 5.3.1 the posterior distribution of the parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}$, for $t \geq I$, is given by

$$\begin{aligned}
&\pi(\boldsymbol{\phi}, \boldsymbol{\sigma} \,|\, \mathcal{D}(t)) \propto g(\mathcal{D}(t) \,|\, \boldsymbol{\phi}, \boldsymbol{\sigma}) f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) f_{\boldsymbol{\sigma}}(\boldsymbol{\sigma}) \\
&\quad = \left[g(C_{1,0}, \ldots, C_{I,0}) \prod_{j=0}^{J-1} \prod_{i=1}^{(I-j-1)\wedge I} \frac{(\phi_j \sigma_j^{-2})^{C_{i,j}\sigma_j^{-2}}}{\Gamma(C_{i,j}\sigma_j^{-2})} C_{i,j+1}^{C_{i,j}\sigma_j^{-2}-1} \exp\left\{-\phi_j \sigma_j^{-2} C_{i,j+1}\right\} \right] \\
&\quad \times \left[\prod_{j=0}^{J-1} \lim_{\gamma_j \to 1} \frac{(f_j(\gamma_j - 1))^{\gamma_j}}{\Gamma(\gamma_j)} \phi_j^{\gamma_j - 1} \exp\left\{-\phi_j f_j(\gamma_j - 1)\right\} \right] \times \left[\prod_{j=0}^{J-1} f_{\sigma_j^2}(\sigma_j^2) \right] \\
&\quad \propto \left[\prod_{j=0}^{J-1} \lim_{\gamma_j \to 1} \phi_j^{\gamma_j - 1 + \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j}\sigma_j^{-2}} \exp\left\{ -\phi_j \left(f_j(\gamma_j - 1) + \sum_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}\sigma_j^{-2} \right) \right\} \right] \\
&\quad \times \left[\prod_{j=0}^{J-1} f_{\sigma_j^2}(\sigma_j^2) \prod_{i=1}^{(I-j-1)\wedge I} \frac{(C_{i,j+1}\sigma_j^{-2})^{C_{i,j}\sigma_j^{-2}}}{\Gamma(C_{i,j}\sigma_j^{-2})} \right] \\
&\quad \propto \left[\prod_{j=0}^{J-1} \phi_j^{\sum_{i=1}^{(t-j-1)\wedge I} C_{i,j}\sigma_j^{-2}} \exp\left\{ -\phi_j \left(\sum_{i=1}^{(t-j-1)\wedge I} C_{i,j+1}\sigma_j^{-2} \right) \right\} \right] \\
&\quad \times \left[\prod_{j=0}^{J-1} f_{\sigma_j}(\sigma_j) \prod_{i=1}^{(I-j-1)\wedge I} \frac{(C_{i,j+1}\sigma_j^{-2})^{C_{i,j}\sigma_j^{-2}}}{\Gamma(C_{i,j}\sigma_j^{-2})} \right].
\end{aligned}$$

From the functional form of $\pi(\boldsymbol{\phi}, \boldsymbol{\sigma} \,|\, \mathcal{D}(t))$ it can be seen that the components ϕ_j of $\boldsymbol{\phi}$ and σ_j of $\boldsymbol{\sigma}$ are independent *a posteriori*, which is a direct consequence of the prior independence. Moreover, since $\pi(\boldsymbol{\phi} \,|\, \boldsymbol{\sigma}, \mathcal{D}(t)) \propto \pi(\boldsymbol{\phi}, \boldsymbol{\sigma} \,|\, \mathcal{D}(t))$, the result in (5.23) follows.

□

Remark 5.3.3. *Given $\boldsymbol{\sigma}$ the model in Model Assumptions 5.3.1 belongs to the family of Bayesian models with conjugate priors that allows for closed form (conditional) posteriors – for details see [Wüthrich, 2015].*

From Theorem 5.3.2, the marginal posterior distribution of the elements of the vector $\boldsymbol{\sigma}$ is given by

$$\pi\big(\sigma_j \,|\, \mathcal{D}(t)\big) \propto h_j(\sigma_j \,|\, \mathcal{D}(t)) = \Gamma(a_j) b_j^{-a_j} f_{\sigma_j}(\sigma_j) \prod_{i=1}^{(I-j-1)\wedge I} \frac{(C_{i,j+1}\sigma_j^{-2})^{C_{i,j}\sigma_j^{-2}}}{\Gamma(C_{i,j}\sigma_j^{-2})}, \tag{5.25}$$

with a_j and b_j defined in (5.24). As seen in the Lemma below as long as some (mild) conditions are satisfied, one can ensure the posterior distribution of $\boldsymbol{\sigma}$ is proper.

Lemma 5.3.4. *For $0 \leq j \leq J-1$ and $t \geq I$ if either $(t-j-1) \wedge I = 1$ or at least one accident year $1 \leq i \leq (t-j-1) \wedge I$ is such that $\frac{C_{i,j+1}}{C_{i,j}} \neq \widehat{f}_j(t)$ then the marginal posterior $\pi(\boldsymbol{\sigma} \,|\, \mathcal{D}(t))$ is integrable, i.e.,*

$$\int_0^{d_j} h_j(\sigma_j \,|\, \mathcal{D}(t)) d\sigma_j < \infty.$$

Proof. See [Peters et al., 2017, Lemma 3.1]. □

Therefore, under Model Assumptions 5.3.1 inference for all the unknown parameters can be performed. It should be noticed, though, that differently from the (conditional) posteriors for ϕ_j (5.23), the posterior for σ_j (5.25) is not recognized as a known distribution. Thus, whenever expectations with respect to the distribution of $\sigma_j \,|\, \mathcal{D}(t)$ need to be calculated one needs to make use of numerical procedures, such as numerical integration or Markov Chain Monte Carlo (MCMC) methods.

Remark 5.3.5. *Note that, comparing Model Assumptions 5.1.1 and 5.3.1 we see that a point estimate for the parameters σ_j from the Bayesian CL model is given by*

$$\widehat{\sigma}_j(t) = \frac{\sqrt{\widehat{s}_j^2(t)}}{\widehat{f}_j(t)}, \ \forall 0 \leq j \leq J-1.$$

5.3.1 MSEP results conditional on σ

Following Model Assumptions 5.3.1 we now discuss how to explicitly calculate the quantities introduced in Section 5.2.

We start with the equivalent of the classic CL factor. From the model structure in (5.22) we define the posterior Bayesian CL factors, given $\boldsymbol{\sigma}$, as

$$\overline{f}_j(t) = \mathbb{E}_{\phi_j}[\phi_j^{-1} \mid \sigma_j, \mathcal{D}(t)], \tag{5.26}$$

which, using the Gamma distribution from (5.23) takes the form

$$\overline{f}_j(t) = \frac{\sum_{k=1}^{(t-j-1)\wedge I} C_{k,j+1}}{\sum_{k=1}^{(t-j-1)\wedge I} C_{k,j}} = \widehat{f}_j(t),$$

where $\widehat{f}_j(t)$ is the classic CL factor estimate.

Following (5.8) we define the conditional ultimate claim predictor

$$\overline{\mathscr{C}}_{i,J}(t) = \mathbb{E}[C_{i,J} \,|\, \boldsymbol{\sigma}, \mathcal{D}(t)] = \mathbb{E}_{\boldsymbol{\phi}}\Big[\mathbb{E}[C_{i,J} \,|\, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathcal{D}(t)] \,\Big|\, \boldsymbol{\sigma}, \mathcal{D}(t)\Big],$$

which can be shown (see [Wüthrich, 2015, Theorem 9.5]) to be equal to

$$\overline{\mathscr{C}}_{i,J}(t) = C_{i,t-i} \prod_{j=t-i}^{J-1} \widehat{f}_j(t) = \widehat{\mathscr{C}}_{i,J}(t), \tag{5.27}$$

where the right hand side is exactly the classic chain ladder estimator of [Mack, 1993]. For this reason we take Model Assumptions 5.3.1 as a distributional model for the classical CL method. Additionally, the conditional reserves defined in (5.8) and (5.13) are also the same as the classic CL ones:

$$\overline{\mathcal{R}}(t) = \sum_{i=1}^{I} \widehat{\mathscr{C}}_{i,J}(t) - C_{i,t-i} = \widehat{\mathcal{R}}(t). \tag{5.28}$$

The importance of equation (5.27) relies on the fact that its right-hand side (the classic CL predictor) does not depend on the parameter vector $\boldsymbol{\sigma}$. In other words, the ultimate claim

predictor based on the Bayesian model from Model Assumptions 5.3.1 conditional on $\boldsymbol{\sigma}$ – which is, in general, a random variable – is a real number (independent of $\boldsymbol{\sigma}$), which justifies the argument used on the calculation of (5.19).

Remark 5.3.6. *Using the notation from the previous sections the parameter vector $\boldsymbol{\sigma}$ plays the role of $\boldsymbol{\theta}$ as the only unknown, since, due to conjugacy properties, $\boldsymbol{\phi}$ can be marginalized analytically.*

For the Bayesian model from Model Assumptions 5.3.1 the msep conditional on $\boldsymbol{\sigma}$ has been derived in [Wüthrich, 2015, Theorem 9.16] as follows

$$\operatorname{msep}_{\overline{\mathrm{CDR}}_i(t+1)\,|\,\boldsymbol{\sigma},\,\mathcal{D}(t)}(0) = \big(\widehat{\mathscr{C}}_{i,J}(t)\big)^2 \left[\left(1+\frac{\overline{\Psi}_{t-i}(t)}{\overline{\beta}_{t-i}(t)}\right)\prod_{j=t-i+1}^{J-1}\big(1+\overline{\beta}_j(t)\overline{\Psi}_j(t)\big)-1\right], \quad (5.29)$$

where

$$\overline{\beta}_j(t)=\frac{C_{t-j,j}}{\sum_{i=1}^{t-j}C_{i,j}} \quad \text{and} \quad \overline{\Psi}_j(t)=\frac{\sigma_j^2}{-\sigma_j^2+\sum_{k=1}^{t-j-1}C_{k,j}}. \quad (5.30)$$

Moreover, the conditional msep has been shown to be finite whenever $\sigma_j^2 < \sum_{k=1}^{t-j-1} C_{k,j}$.

The aggregated conditional msep for $\overline{\mathrm{CDR}}(t+1)=\sum_{i=1}^{I}\overline{\mathrm{CDR}}_i(t+1)$ is also derived in [Wüthrich, 2015, Theorem 9.16], and given by

$$\begin{aligned}\operatorname{msep}_{\overline{\mathrm{CDR}}(t+1)\,|\,\boldsymbol{\sigma},\,\mathcal{D}(t)}(0) &= \sum_{i=t-J+1}^{I}\operatorname{msep}_{\overline{\mathrm{CDR}}_i(t+1)\,|\,\boldsymbol{\sigma},\,\mathcal{D}(t)}(0)\\ &+2\sum_{t-J+1\le i<k\le I}\sum \widehat{\mathscr{C}}_{i,J}(t)\widehat{\mathscr{C}}_{k,J}(t)\left[\big(1+\overline{\Psi}_{t-i}(t)\big)\prod_{j=t-i+1}^{J-1}\Big(1+\overline{\beta}_j(t)\overline{\Psi}_j(t)\Big)-1\right]. \quad (5.31)\end{aligned}$$

Remark 5.3.7. *The assumption that $\sigma_j^2 < \sum_{k=1}^{t-j-1} C_{k,j}$ is made in order to guarantee the conditional msep is finite and we enforce this assumption to hold for all the examples presented in this work.*

5.3.2 Marginalized MSEP results

The results in the previous section are based on derivations presented in [Merz and Wüthrich, 2015] and [Wüthrich, 2015] where the parameter vector $\boldsymbol{\sigma}$ is assumed to be known. In this section we study the impact of the uncertainty in $\boldsymbol{\sigma}$ over the mean and variance of $\mathscr{C}_{i,J}(t+1)\,|\,\mathcal{D}(t)$ in light of Model Assumptions 5.3.1, which can be seen as a fully Bayesian version of the models previously mentioned.

In order to have well defined posterior distributions for $\boldsymbol{\sigma}$, through this section we follow Lemma 5.3.4 and assume that, for all development years $0 \le j \le J-1$ and $t \ge I$, we have $(t-j-1)\wedge I = 1$ or at least one accident year $1 \le i \le (t-j-1)\wedge I$ is such that $\frac{C_{i,j+1}}{C_{i,j}} \neq \widehat{f}_j(t)$. For all the results presented this (very mild) assumption is satisfied.

Lemma 5.3.8. *The ultimate claim estimator under the marginalized model is equal to the classic chain ladder predictor, i.e., $\mathscr{C}_{i,J}(t)=\mathbb{E}[C_{i,J}\,|\,\mathcal{D}(t)]=\widehat{\mathscr{C}}_{i,J}(t)$.*

Proof. Due to the posterior independence of the elements of $\boldsymbol{\phi}$ and the fact that $\overline{\mathscr{C}}_{i,J}(t) = \widehat{\mathscr{C}}_{i,J}(t)$ does not depend on $\boldsymbol{\sigma}$ we have

$$\begin{aligned}
\mathscr{C}_{i,J}(t) &= \mathbb{E}[C_{i,J} \,|\, \mathcal{D}(t)] \\
&= \mathbb{E}_{(\boldsymbol{\phi},\boldsymbol{\sigma})}\big[\mathbb{E}[C_{i,J} \,|\, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathcal{D}(t)] \,|\, \mathcal{D}(t)\big] \\
&= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\mathbb{E}_{\boldsymbol{\phi}}\big[\mathbb{E}[C_{i,J} \,|\, \boldsymbol{\phi}, \boldsymbol{\sigma}, \mathcal{D}(t)] \,|\, \boldsymbol{\sigma}, \mathcal{D}(t)\big] \,|\, \mathcal{D}(t)\Big] \\
&= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\mathbb{E}_{\boldsymbol{\phi}}\big[C_{i,\,t-i} \prod_{j=t-i}^{J-1} \phi_j^{-1} \,|\, \boldsymbol{\sigma}, \mathcal{D}(t)\big] \,|\, \mathcal{D}(t)\Big] \\
&= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\overline{\mathscr{C}}_{i,J}(t) \,|\, \mathcal{D}(t)\Big] = \widehat{\mathscr{C}}_{i,J}(t).
\end{aligned}$$

□

Proposition 5.3.9. *The msep in the marginalized model is equal to the posterior expectation of the msep in the conditional model, i.e.,*

$$\begin{aligned}
msep_{CDR(t+1)\,|\,\mathcal{D}(t)}(0) &= Var\Big(\sum_{i=1}^{I} \mathscr{C}_{i,J}(t+1) \,\Big|\, \mathcal{D}(t)\Big) \\
&= \mathbb{E}_{\boldsymbol{\sigma}}[msep_{\overline{CDR}(t+1)\,|\,\boldsymbol{\sigma},\,\mathcal{D}(t)}(0) \,|\, \mathcal{D}(t)]. \qquad (5.32)
\end{aligned}$$

Proof. From the law of total variance we have that

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^{I} \mathscr{C}_{i,J}(t+1) \,\Big|\, \mathcal{D}(t)\Big) &= \mathrm{Var}_{\boldsymbol{\sigma}}\Big(\mathbb{E}\big[\sum_{i=1}^{I} \mathscr{C}_{i,J}(t+1) \,|\, \mathcal{D}(t), \boldsymbol{\sigma}\big] \,\Big|\, \mathcal{D}(t)\Big) \\
&\quad + \mathbb{E}_{\boldsymbol{\sigma}}\Big[\mathrm{Var}\big(\sum_{i=1}^{I} \mathscr{C}_{i,J}(t+1) \,|\, \mathcal{D}(t), \boldsymbol{\sigma}\big) \,\Big|\, \mathcal{D}(t)\Big] \\
&= \mathbb{E}_{\boldsymbol{\sigma}}\Big[\mathrm{Var}\big(\sum_{i=1}^{I} \mathscr{C}_{i,J}(t+1) \,|\, \mathcal{D}(t), \boldsymbol{\sigma}\big) \,\Big|\, \mathcal{D}(t)\Big],
\end{aligned}$$

and the last equality follows from Lemma 5.3.8 and the fact that $\mathbb{E}[\widehat{\mathscr{C}}_{i,J}(t+1) \,|\, \mathcal{D}(t), \boldsymbol{\sigma}] = \widehat{\mathscr{C}}_{i,J}(t)$ is independent of $\boldsymbol{\sigma}$. □

Remark 5.3.10. *Following the conditions required for finiteness of the conditional msep, in the unconditional case, one can see that* $msep_{CDR(t+1)\,|\,\mathcal{D}(t)}(0) < +\infty$ *whenever* $\sum_{k=1}^{t-j-1} C_{k,j} > d_j^2$. *Furthermore, we note that this condition can be controlled during the model specification.*

5.3.3 Statistical model of PY risk in the SST

Note that the distributional models derived in Sections 5.3.1 and 5.3.2 are rather complex. To maintain some degree of tractability the overall PY uncertainty distribution is usually approximated by a log-normal distribution via a moment matching procedure.

5.3.3.1 Conditional PY model

As discussed in Section 5.2.3, when modelling the risk of PY claims we work with the random variables $\overline{Z}_{PY}$, defined in (5.16). Due to their relationship with the conditional CDR, see (5.9)

and (5.10) and the results discussed in Section 5.3.1, we can use the derived properties of these random variables to construct the model being used for $\overline{Z}_{PY}$.

The conditional mean (see (5.9), (5.10) and (5.28)) and variance (see (5.12) and (5.31)) of the random variable $\overline{Z}_{PY}$ are as follows

$$\mathbb{E}[\overline{Z}_{PY} \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t)] = \widehat{\mathcal{R}}(t), \tag{5.33}$$

$$\mathrm{Var}(\overline{Z}_{PY} \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t)) = \mathrm{msep}_{\overline{\mathrm{CDR}}(t+1) \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t)}(0). \tag{5.34}$$

Given mean and variance, we make the following approximation.

Model Assumptions 5.3.11 (Conditional log-normal approximation)**.** *We assume that*

$$\overline{Z}_{PY} \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t) \sim LN\Big(\overline{\mu}_{PY},\, \overline{\sigma}^2_{PY}\Big),$$

with $\overline{\sigma}^2_{PY} = \log\left(\frac{msep_{\overline{CDR}(t+1) \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t)}(0)}{\widehat{\mathcal{R}}(t)^2} + 1\right)$ *and* $\overline{\mu}_{PY} = \log\big(\widehat{\mathcal{R}}(t)\big) - \frac{\overline{\sigma}^2_{PY}}{2}$.

Although the distribution of $\overline{Z}_{PY} \,|\, \boldsymbol{\sigma},\, \mathcal{D}(t)$ under Model Assumptions 5.3.1 can not be described analytically it is simple to simulate from it. To test the approximation of Model Assumptions 5.3.11 we simulate its distribution under the Bayesian gamma-gamma model (with fixed $\boldsymbol{\sigma}$) and compare it against the log-normal approximation proposed. For the hyper-parameters presented in Table 6.5 (and calculated in Section 6.3) the quantile-quantile plot of the approximation is presented in Figure 5.1. For all the LoBs we see the log-normal is a sensible approximation to the original model assumptions. Note that although the parameters used for the comparison are based on the marginalized model, Figures 6.5 and 6.6 show that they are "representative" values for the distributions of $\overline{\mu}_{PY}$ and $\overline{\sigma}_{PY}$.

5.3.3.2 Marginalized PY model

As an alternative to the conditional Model Assumptions 5.3.11 we use the moments of $Z_{PY} \,|\, \mathcal{D}(t)$ calculated in Lemma 5.3.8 and Proposition 5.3.9 and then approximate its distribution. Note that due to the intractability of the distribution of $\boldsymbol{\sigma} \,|\, \mathcal{D}(t)$ the variance term defined in (5.32) can only be calculated numerically, for example, via MCMC.

Model Assumptions 5.3.12 (Marginalized log-normal approximation)**.** *We assume that*

$$Z_{PY} \,|\, \mathcal{D}(t) \sim LN\Big(\mu_{PY},\, \sigma^2_{PY}\Big)$$

with $\sigma^2_{PY} = \log\left(\frac{msep_{CDR(t+1) \,|\, \mathcal{D}(t)}(0)}{\widehat{\mathcal{R}}(t)^2} + 1\right)$ *and* $\mu_{PY} = \log\big(\widehat{\mathcal{R}}(t)\big) - \frac{\sigma^2_{PY}}{2}$.

The same comparison based on the quantile-quantile plot of Figure 5.1 can be performed for the marginalized model and the results are presented in Figure 5.2. Once again, the log-normal model presents a viable alternative to the originally postulated gamma-gamma model, even though for Motor Hull, Property and Others the right tail of the log-normal distribution is slightly heavier.

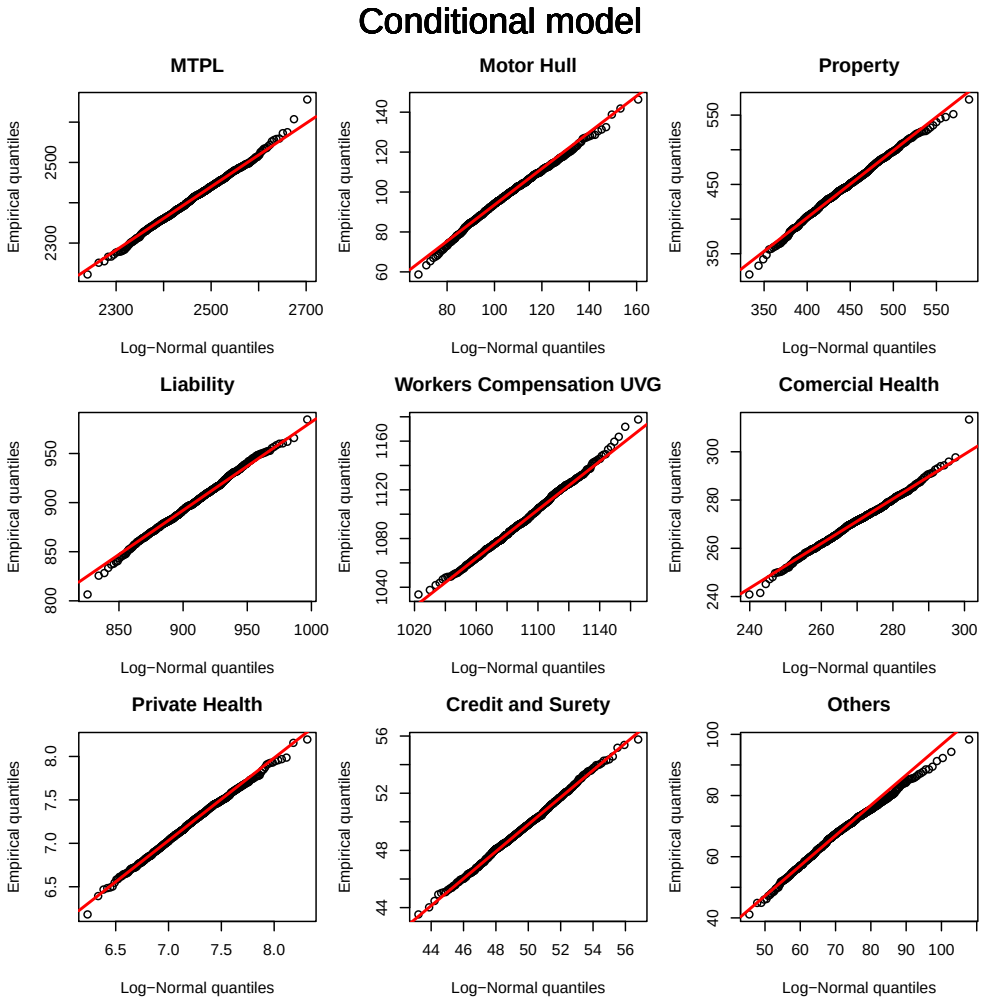

Figure 5.1: Quantile-Quantile plots, using the data from Figure 6.1, for the different LoBs comparing (vertical axis) the empirical distribution of $\overline{Z}_{PY} \,|\, \boldsymbol{\sigma}$, $\mathcal{D}(t)$ based on Model Assumptions 5.3.1 and (horizontal axis) the log-normal approximation from Model Assumptions 5.3.11.

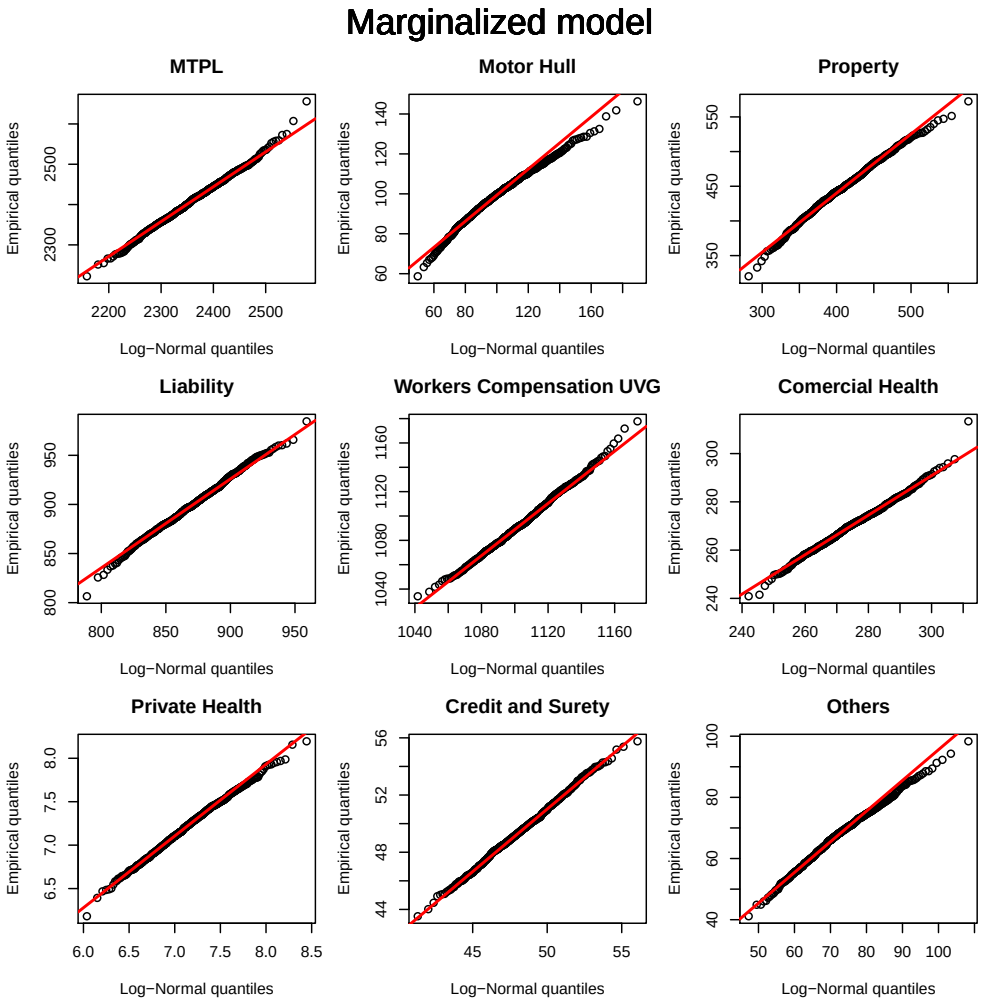

Figure 5.2: Quantile-Quantile plots, using the data from Figure 6.1, for the different LoBs comparing (vertical axis) the empirical distribution of $Z_{PY} \,|\, \mathcal{D}(t)$ based on Model Assumptions 5.3.1 and (horizontal axis) the log-normal approximation from Model Assumptions 5.3.12 and using posterior samples as in Figures 6.5 and 6.6.

5.4 Modelling of individual LoBs CY claims

Model Assumptions 5.3.1 do not assume any specific distribution for $C_{t+1,0}$, the CY claims. These claims are treated differently from PY claims and the models used for these claim random variables are explained in Sections 5.4.1 and 5.4.2, below. Throughout this section we denote by $\lambda_{CY} = \lambda_{CY,s} + \lambda_{CY,l}$ the expected number of CY claims over the next year, which is the sum of the CY small claims ($\lambda_{CY,s}$) and the CY large claims ($\lambda_{CY,l}$).

5.4.1 Modelling of small CY claims

As mentioned in the SST Technical Document [FINMA, 2007, Section 4.4.7], *the SST does not make any explicit assumption about the distribution of individual claims; instead, the annual claims expenses are only represented with their expected value and variance.* More precisely, in [FINMA, 2007, Section 8.4.5.2] the distribution of the premium risk, $Z_{CY,s}$ is assumed to be such that the square of its coefficient of variation (standard deviation divided by its mean) is equal to

$$\mathrm{CoVa}^2(Z_{CY,s} \,|\, \mathcal{D}(t)) = A_1 + \frac{A_2 + 1}{\lambda_{CY,s}}, \tag{5.35}$$

where the constants A_1 and A_2 are provided by the regulatory authority (under the names of *parameter uncertainty* and *random fluctuation*, respectively). Its values for the 2015 solvency test are found in [FINMA, 2016]. The (known) constant $\lambda_{CY,s}$ denote the expected number of small claims over the next calendar year.

In order to fully specify the model for CY small claims one also needs to decide on the mean of the variable $Z_{CY,s} \,|\, \mathcal{D}(t)$, but we postpone a detailed discussion on this point until Section 6.3.2, where we also present the value of $\lambda_{CY,s}$.

Model Assumptions 5.4.1 (Distribution of CY small claims)**.** *For known constants v, $r_s > 0$ and $\mathbb{E}[Z_{CY,s} \,|\, \mathcal{D}(t)]$ we set*

$$Z_{CY,s} \,|\, \mathcal{D}(t) \sim LN\Big(\mu_{CY,s},\, \sigma^2_{CY,s}\Big)$$

with $\sigma^2_{CY,s} = \log\left(A_1 + \dfrac{A_2 + 1}{\lambda_{CY,s}} + 1\right)$ *and* $\mu_{CY,s} = \log(\mathbb{E}[Z_{CY,s} \,|\, \mathcal{D}(t)]) - \dfrac{\sigma^2_{CY,s}}{2}$.

5.4.2 Distribution of large CY claims

In the SST (see [FINMA, 2007, Section 4.4.8]) large CY claims are split into two groups. The first group encompasses individual claims with a large claim amount, which includes, as exemplified in [FINMA, 2007, Section 4.4.8] fire in a factory building. The second group of large claims are those triggered by the same event (e.g. a hailstorm) but with many simultaneous claims. These types of claims are likely to affect all market participants.

For each risk trigger, CY large claims are required to be modelled as a compound Poisson process with Pareto severities, i.e.,

$$Z_{CY,l} = \sum_{k=1}^{N} Y_k, \tag{5.36}$$

where $N \sim Poi(\lambda)$ is the number of large claims in LoB under consideration and $Y_k \sim Pareto(\beta, \alpha_\beta)$ model the intensity of large claims. It is assumed in the SST that claims are i.i.d. within the same risk trigger and independent of all Z_{PY} and $Z_{CY,s}$.

As a notational remark, if Z follows a Compound Poisson - Pareto model as a shorthand notation we write $Z \sim$ CP-P(λ, β, α), with the same parameter interpretation as in (5.36).

5.4.2.1 SST model for cumulated claims

In this section we discuss the modelling of cumulated claims (those triggered by a market-wide event) which are modelled as an event that impacts the whole market and then scales down to an individual insurance company through its market share. In particular we present the modelling approach used in the (1) Motor Hull LoB due to hail events and (2) Workers Compensation (UVG) LoB due to a market-wide large accidents.

In both cases market-wide parameters for a compound Poisson model with Pareto intensity have been calculated by the regulator, based on an extensive claims data set. The aggregated market-wide loss is given by

$$Z_{mkt} = \sum_{k=1}^{N_{mkt}} Y_{k,\,mkt} \sim \text{CP-P}(\lambda_{mkt},\, \beta_{mkt},\, \alpha_{mkt}).$$

The corresponding market-wide parameter values are found in [FINMA, 2016].

Denoting by β the company's threshold after which losses are classified as large and m its market share in the ℓ-th LoB, to be consistent with its assumption the company should model market-wide large events as events above the threshold of

$$\beta_{mkt,\,comp} = \frac{\beta}{m}.$$

Then, the market-wide total loss (viewed from the specific company in consideration) is defined as

$$Z_{mkt,\,comp} = \sum_{k=1}^{N_{mkt,\,comp}} Y_{k,\,mkt,\,comp} \sim \text{CP-P}(\lambda_{mkt,\,comp},\, \beta_{mkt,\,comp},\, \alpha_{mkt}),$$

from which it is easy to see the only unknown parameter is $\lambda_{mkt,\,comp}$, since in the SST the Pareto parameter α_{mkt} is kept the same. This frequency parameter is chosen such that the company's view of the market-wide events is equivalent to the suggested market-wide process. In other words, if $\overset{d}{=}$ denotes equality in distribution, $\lambda_{mkt,\,comp}$ is chosen in order to have

$$Z_{mkt} \overset{d}{=} Z_{mkt,\,comp},$$

from which it is then easy to see that $\lambda_{mkt} = \mathbb{P}[Y_{k,\,mkt,\,comp} > \beta_{mkt}]\lambda_{mkt,\,comp}$ hence

$$\lambda_{mkt,\,comp} = \lambda_{mkt} \left(\frac{\beta/m}{\beta_{mkt}} \right)^{-\alpha_{mkt}}.$$

Therefore, from the company's point of view, its own large claims are modelled as

$$Z_{comp} \overset{d}{=} m \times Z_{mkt,\,comp} = \sum_{k=1}^{N_{mkt,\,comp}} m \times Y_{k,\,mkt,\,comp},$$

which implies that

$$Z_{comp} \sim \text{CP-P}(\lambda_{mkt,\,comp},\, \beta,\, \alpha_{mkt}).$$

Since the distribution of Z_{comp} is not known in closed form we will assume it is sufficiently well approximated by a single Pareto distribution, with the same lower truncation point β and mean as Z_{comp}, i.e.,

$$\mathbb{E}[Z_{comp}] = \lambda_{mkt} \left(\frac{\beta/m}{\beta_{mkt}}\right)^{-\alpha_{mkt}} \frac{\alpha_{mkt}}{\alpha_{mkt}-1}\beta. \tag{5.37}$$

As in the SST Technical Document [FINMA, 2007], an upper bound γ (provided by the regulator) is included after matching the lower bound.

Model Assumptions 5.4.2 (Marginal distribution of cumulated claims)**.** *For α_{mkt}, β_{mkt} and γ provided by the regulator in [FINMA, 2016], $\beta \in \{1,\, 5\}$, $m \in (0, 1)$,*

$$Z_{CY,l} \sim \textit{Pareto}\left(\lambda_{mkt}\left(\frac{\beta/m}{\beta_{mkt}}\right)^{-\alpha_{mkt}} \beta,\, \alpha_{mkt},\, \gamma\right),$$

where $\textit{Pareto}(\beta, \alpha, \gamma)$ denotes a Pareto distribution defined in $[\beta,\, \gamma]$ and tail index α.

Remark 5.4.3. *The reader should note that for large CY claims there is no parameter uncertainty, since both λ_{mkt}, α_{mkt} and γ are given by the regulator, the market share, m can be perfectly calculated and β is chosen by the company.*

5.4.2.2 SST model for individual claims

For individual large events, the SST provides p_1, the probability of observing losses larger than CHF 1 million and standard values for α_β, for $\beta = 1$ and $\beta = 5$ (see Table [FINMA, 2016]). Since the probability of large claims provided by the SST is based on a lower threshold of CHF 1 million, a thinning process of the CP-P has to be done if the company decides to use $\beta = 5$.

Following the same procedure presented in Section 5.4.2.1 we can see that

$$Z_{CY,l} \sim \text{CP-P}(\lambda_\beta,\, \beta,\, \alpha_\beta),$$

with an expected number of claims larger than β equal to

$$\lambda_\beta = \lambda_{CY,l} = p_1 \lambda_{CY} \left(\frac{\beta}{1}\right)^{-\alpha_\beta}, \tag{5.38}$$

where λ_{CY} denotes the expected total number of CY claims in the ℓ-th LoB.

As in Section 5.4.2.1 the distribution of $Z_{CY,l} \,|\, \mathcal{D}(t)$ is approximated by a single Pareto, with the same mean and Pareto index α_β. Likewise, an upper bound γ is also introduced.

Model Assumptions 5.4.4 (Marginal distribution of large individual claims)**.** *For α_β, p_1 and γ provided by the regulator in [FINMA, 2016], $\beta \in \{1,\, 5\}$ and $\lambda_{CY} > 0$,*

$$Z_{CY,l} \,|\, \mathcal{D}(t) \sim \textit{Pareto}\left(p_1 \lambda_{CY} \beta^{1-\alpha_\beta},\, \alpha_\beta,\, \gamma\right).$$

5.5 Joint distribution of PY and CY claims

Although the SST does not assume any parametric form for the joint distribution of $\boldsymbol{Z} \,|\, \mathcal{D}(t)$ or $\overline{\boldsymbol{Z}} \,|\, \mathcal{D}(t)$ (defined in (5.18) and (5.17), respectively) it is required that a pre-specified *correlation matrix* Λ is used (see [FINMA, 2016]). In this section we discuss how to use the conditional and marginalized models to define a joint distribution satisfying this correlation.

It is important to notice, though, that the SST correlation matrix may not be attainable for some joint distributions, as discussed in Section 2.1.3 in the case of log-normal marginals.

It should be noted that, since in the SST the CY large claims are assumed to be independent from all the other risks, the correlation matrix of $(\boldsymbol{Z}_{PY},\ \boldsymbol{Z}_{CY,s},\ \boldsymbol{Z}_{CY,l}) \,|\, \mathcal{D}(t)$ is essentially a correlation matrix between $(\boldsymbol{Z}_{PY},\ \boldsymbol{Z}_{CY,s}) \,|\, \mathcal{D}(t)$ and the same is true also for the conditional model.

Regardless of assuming a conditional or a marginalized model, SST's correlation matrix Λ should be such that, for $i, j = 1, \dots, 2L + P$,

$$\Lambda_{i,j} = \mathrm{Corr}(Z_i,\ Z_j \,|\, \mathcal{D}(t)) = \mathrm{Corr}(\overline{Z}_i,\ \overline{Z}_j \,|\, \mathcal{D}(t)).$$

Remark 5.5.1. *In the conditional model we need to "integrate out" the parameter uncertainty, otherwise the (conditional) correlation would be dependent on an unknown parameter and could not be matched with the numbers provided by the SST.*

5.5.1 Conditional joint model

Under Model Assumptions 5.3.11, 5.4.1, 5.4.2 and 5.4.4 our interest lies on modelling the joint behaviour of the vector $\overline{\boldsymbol{Z}} \,|\, \boldsymbol{\sigma},\ \mathcal{D}(t)$. Under Model Assumptions 5.3.1 it can be shown that the required conditional independence between $\boldsymbol{Z}_{CY,l}$ and $(\boldsymbol{Z}_{PY},\ \boldsymbol{Z}_{CY,s})$ given $\mathcal{D}(t)$ is equivalent to the conditional independence between $\boldsymbol{Z}_{CY,l}$ and $(\boldsymbol{Z}_{PY},\ \boldsymbol{Z}_{CY,s})$ given $\mathcal{D}(t),\ \boldsymbol{\sigma}$.

Moreover, since all the marginal conditional distributions are assumed to be log-normal, following (5.18) and (5.17) the notation can be further simplified to

$$\overline{Z}_i \,|\, \boldsymbol{\sigma},\ \mathcal{D}(t) \sim \mathrm{LN}(\overline{m}_i(\boldsymbol{\sigma}),\ \overline{V}_i(\boldsymbol{\sigma})),\ \text{for } i = 1, \dots, 2L, \tag{5.39}$$

with $\overline{m}_i(\boldsymbol{\sigma})$, and $\overline{V}_i(\boldsymbol{\sigma})$ defined on Model Assumptions 5.3.11 and 5.4.1. For example, for $i = L + 1$, $\overline{m}_i(\boldsymbol{\sigma}) = \mu^{(1)}_{CY,s}$, defined on Model Assumptions 5.4.1.

We are now ready to define the joint conditional model to be used.

Model Assumptions 5.5.2 (Conditional joint model)**.** *Based on Model Assumptions 5.3.11 and 5.4.1 we link the marginals of the conditional model through a Gaussian copula with correlation matrix $\overline{\Omega}$. More formally, given $\mathcal{D}(t)$ and $\boldsymbol{\sigma}$, the joint distribution of $\overline{\boldsymbol{Z}}$ is given by*

$$F_{\overline{\boldsymbol{Z}}}(\overline{z}_1, \dots, \overline{z}_{2L};\ \overline{\Omega} \,|\, \mathcal{D}(t),\ \boldsymbol{\sigma}) = C\Big(F_{\overline{Z}_1}(\overline{z}_1 \,|\, \mathcal{D}(t),\ \boldsymbol{\sigma}), \dots, F_{\overline{Z}_{2L}}(\overline{z}_{2L} \,|\, \mathcal{D}(t),\ \boldsymbol{\sigma});\ \overline{\Omega}\Big),$$

where $F_{\overline{Z}_i}(\,\cdot \,|\, \mathcal{D}(t),\ \boldsymbol{\sigma})$ denotes the conditional distribution of $\overline{Z}_i \,|\, \mathcal{D}(t),\ \boldsymbol{\sigma}$ defined in (5.39) and $C(\,\cdot\,; \overline{\Omega})$ is the Gaussian copula with correlation matrix $\overline{\Omega}$.

Remark 5.5.3. *In this section the parameter matrix $\overline{\Omega}$ should be understood as a deterministic variable, differently from $\boldsymbol{\sigma}$ and $\boldsymbol{\phi}$. For this reason we do not include it on the right hand side of the conditioning bar. Instead, whenever $\overline{\Omega}$ needs to be explicitly written, we include it on the left hand side of the bar, separated by the function (or functional, for expectations) arguments by a semicolon.*

In order to match SST's correlation matrix Λ, under Model Assumptions 5.3.1 and 5.4.1, the following equation needs to be solved with respect to $\overline{\Omega}$:

$$\Lambda_{i,j} = \mathrm{Corr}(\overline{Z}_i,\, \overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t)). \tag{5.40}$$

To compute the right hand side of the equation above we first notice that

$$\mathrm{Cov}(\overline{Z}_i,\, \overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t)) = \mathbb{E}[\overline{Z}_i\overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t)] - \mathbb{E}[\overline{Z}_i\,|\,\mathcal{D}(t)]\mathbb{E}[\overline{Z}_j\,|\,\mathcal{D}(t)],$$

where, from (5.33) and the discussion in Section 5.4.1

$$\begin{aligned}\mathbb{E}[\overline{Z}_i\,|\,\mathcal{D}(t)] &= \mathbb{E}_{\boldsymbol{\sigma}}[\mathbb{E}[\overline{Z}_i\,|\,\mathcal{D}(t),\, \boldsymbol{\sigma}]\,|\,\mathcal{D}(t)] \\ &= \mathbb{E}_{\boldsymbol{\sigma}}[\overline{m}_i\,|\,\mathcal{D}(t)] = \begin{cases} \widehat{\mathcal{R}}^{(i)}(t), \text{ if } 1 \le i \le L \\ \mathbb{E}[Z_{CY,s}^{(i-L)}\,|\,\mathcal{D}(t)], \text{ if } L+1 \le i \le 2L \end{cases}\end{aligned}$$

and from (2.4), Section 2.1.3,

$$\begin{aligned}\mathbb{E}[\overline{Z}_i\overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t)] &= \mathbb{E}_{\boldsymbol{\sigma}}[\mathbb{E}[\overline{Z}_i\overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t),\, \boldsymbol{\sigma}]\,|\,\mathcal{D}(t)] \\ &= \mathbb{E}_{\boldsymbol{\sigma}}\left[\exp\left\{\overline{m}_i + \frac{\overline{V}_i^2 + 2\overline{V}_i\overline{\omega}_{i,j}\overline{V}_j + \overline{V}_j^2}{2} + \overline{m}_j\right\}\,\Big|\,\mathcal{D}(t)\right].\end{aligned}$$

Therefore, to satisfy (5.40) $\overline{\Omega}_{i,j}$ needs to be chosen such that the following implicit relationship (which can be solved through any univariate root search algorithm) holds:

$$\Lambda_{i,j}\sqrt{\mathrm{Var}(\overline{Z}_i\,|\,\mathcal{D}(t))\mathrm{Var}(\overline{Z}_j\,|\,\mathcal{D}(t))} + \mathbb{E}[\overline{Z}_i\,|\,\mathcal{D}(t)]\mathbb{E}[\overline{Z}_j\,|\,\mathcal{D}(t)] - \mathbb{E}[\overline{Z}_i\overline{Z}_j;\, \overline{\Omega}\,|\,\mathcal{D}(t)] = 0.$$

5.5.2 Marginalized joint model

Similarly to Section 5.5.1 in this section we will fully characterize the joint distribution of $\boldsymbol{Z}\,|\,\mathcal{D}(t)$ under Model Assumptions 5.3.12, 5.4.1, 5.4.2 and 5.4.4.

From these assumptions we define the following notation:

$$Z_i\,|\,\mathcal{D}(t) \sim \mathrm{LN}(m_i,\, V_i), \text{ for } i = 1, \dots, 2L. \tag{5.41}$$

Model Assumptions 5.5.4 (Marginalized joint model)**.** *Based on Model Assumptions 5.3.12 and 5.4.1 we link the marginal distributions of the marginalized distributions through a Gaussian copula with correlation matrix Ω. More formally, given $\mathcal{D}(t)$, the joint distribution of $\boldsymbol{Z}$ is given by*

$$F_{\boldsymbol{Z}}(z_1, \dots, z_{2L};\, \Omega\,|\,\mathcal{D}(t)) = C\Big(F_{Z_1}(z_1\,|\,\mathcal{D}(t)), \dots, F_{Z_{2L}}(z_{2L}\,|\,\mathcal{D}(t));\, \Omega\Big),$$

where F_{Z_i} denotes the conditional distribution of $Z_i\,|\,\mathcal{D}(t)$ defined in (5.41) and $C(\,\cdot\,;\Omega)$ is the Gaussian copula with correlation matrix Ω.

In order to match SST's correlation matrix, in the joint marginalized model the Gaussian copula correlation Ω is chosen such that (see Equation (2.5))

$$\Lambda_{i,j} = \frac{\exp\{V_i \omega_{i,j} V_j\} - 1}{\left[(e^{V_i^2} - 1)(e^{V_j^2} - 1)\right]^{1/2}}.$$

Chapter 6

Risk allocation under the Swiss Solvency Test

In light of the multivariate distributions based on the marginalized and conditional models from Chapter 5, in this chapter we discuss two SMC-based algorithms for efficient capital allocation. One the one hand, as the algorithm presented in Chapter 4 does not (directly) account for parameter uncertainty it can be straightforwardly applied to the marginalized model. On the other hand, since the expectations involved in the capital allocations for the conditional model are taken with respect to an intractable density, we develop a pseudo-marginal version of the algorithm from Chapter 4 (without the transformation to the unit cube). We also present the balance sheet of a synthetic non-life insurance company from where we generate claims triangles to be used in the inferential procedure. Details of the SMC algorithms and some results conclude the chapter.

Parts of this chapter are based on the working paper [Peters et al., 2016b].

6.1 Risk allocation for the SST

In this section we follow the Euler allocation principle from Section 2.2 and discuss how the risk capital that is held by an insurance company can be split into different risk triggers. As stochastic models for these risks involve a set of unknown parameters, we present an allocation procedure for a marginalized model (which arises when the parameter uncertainty is resolved beforehand) and a conditional model (which is still dependent on an unknown parameter).

Our interest is to split the given overall risk capital into its different LoBs and within each LoB we are also interested in differentiating the impact on the capital of claims from previous accident years and new premium liability. The formal model construction for these risks is presented in Sections 5.2 to 5.5.

As in Chapter 5 let us assume we have two vectors characterizing these risks, namely $\overline{\boldsymbol{Z}} = (\overline{Z}_1, \dots, \overline{Z}_d)$ and $\boldsymbol{Z} = (Z_1, \dots, Z_d)$. In the following sections we discuss the construction of two models: (1) a marginalized one, for the random variable $\boldsymbol{Z}$, where the parameter uncertainty has been previously marginalized; and (2) a conditional one, where risks $\overline{\boldsymbol{Z}}$ are modelled given

some parameter $\boldsymbol{\theta}$. Both models assume knowledge of all claims payments up to the current accounting year and aim at forecasting the outstanding liability (see Section 5.2).

These two models (marginalized and conditional) are defined through the probability density functions (p.d.f.'s), $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $f_{\overline{\boldsymbol{Z}}|\boldsymbol{\theta}}(\overline{\boldsymbol{z}}\,|\,\boldsymbol{\theta})$, respectively. For the conditional model we follow the Bayesian paradigm and assign a prior distribution with p.d.f. $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ to the unknown parameter $\boldsymbol{\theta}$. This is then combined with the likelihood $f_{\overline{\boldsymbol{Z}}|\boldsymbol{\theta}}(\overline{\boldsymbol{z}}\,|\,\boldsymbol{\theta})$ to construct $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$, as seen in (5.6) and (5.7).

Under the marginalized model we define $S = \sum_{i=1}^{d} Z_i$ as the company's overall claims risk. The SST requires the total capital to be calculated as the 99% expected shortfall of S, given by

$$\rho(S) = \mathbb{E}[S\,|\,S \geq \mathrm{VaR}_{99\%}(S)]. \tag{6.1}$$

In turn, the Euler allocation principle states that the contribution of each component Z_i to the capital in (6.1) is given by

$$\mathcal{A}_i = \mathbb{E}[Z_i\,|\,S \geq \mathrm{VaR}_{99\%}(S)], \quad \forall i = 1, \ldots, d. \tag{6.2}$$

The allocations for the conditional model follow the same structure, with Z_i and S replaced, respectively, by $\overline{Z}_i$ and $\overline{S}$ in (6.2) and reads as

$$\overline{\mathcal{A}}_i = \mathbb{E}[\overline{Z}_i\,|\,\overline{S} \geq \mathrm{VaR}_{99\%}(\overline{S})], \quad \forall i = 1, \ldots, d, \tag{6.3}$$

with $\overline{S} = \sum_{i=1}^{d} \overline{Z}_i$. For the models discussed in Chapter 5 the density of $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ is not known in closed form, adding one more layer of complexity to the proposed method.

Remark 6.1.1. *Due to approximations performed on the models introduced in Chapter 5 the allocations derived from the marginalized and conditional models do not match perfectly and it is out of the scope of this work to discuss issues on model selection.*

Although computing $\mathcal{A}_i$ and $\overline{\mathcal{A}}_i$ is a *static* problem, for the sake of transforming the Monte Carlo estimation into an efficient computational framework, we embed the calculation of these quantities into a sequential procedure, where at each step we solve a simpler problem, in a strategy similar to the one adopted in Chapter 4.

6.2 SMC samplers and capital allocation

For the marginalized and conditional models presented in Sections 5.2 to 5.5 the marginal contributions in (6.2) and (6.3) cannot be calculated in analytic form for a generic model, so a simulation technique needs to be employed. Similarly to Chapter 4, in the sequel we discuss how to design SMC algorithms to compute the capital allocations.

6.2.1 Allocations for the marginalized model

For a generic random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ with known marginal densities and distribution functions, respectively $f_{Z_i}(z_i)$ and $F_{Z_i}(z_i)$, and copula density $c(u_1, \ldots, u_d)$, due to Sklar's

theorem (see Theorem 2.1.2) the joint density of $\boldsymbol{Z}$ can be written as

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = c(\boldsymbol{u}) \prod_{i=1}^{d} f_{Z_i}(z_i),$$

where $\boldsymbol{u} = (u_1, \ldots, u_d)$ and $u_i = F_{Z_i}(x_i)$. In order to approximate the marginal risk contributions $\mathcal{A}_i$ from (6.2) we can use samples from the distribution

$$\pi(\boldsymbol{z}) = f_{\boldsymbol{Z}}(\boldsymbol{z} \,|\, \boldsymbol{z} \in \mathcal{G}_{\boldsymbol{Z}}) = \frac{f_{\boldsymbol{Z}}(\boldsymbol{z}) \mathbb{1}_{\mathcal{G}_{\boldsymbol{Z}}}(\boldsymbol{z})}{\mathbb{P}[\boldsymbol{Z} \in \mathcal{G}_{\boldsymbol{Z}}]}, \tag{6.4}$$

where the set $\mathcal{G}_{\boldsymbol{Z}} = \mathcal{G}_{\boldsymbol{Z}}(B)$ is defined, for $B = \mathrm{VaR}_{99\%}(S)$, as

$$\mathcal{G}_{\boldsymbol{Z}} = \left\{ \boldsymbol{z} \in \mathbb{R}^d \,:\, \sum_{i=1}^{d} z_i \geq B \right\}, \tag{6.5}$$

and the indicator function $\mathbb{1}_{\mathcal{G}_{\boldsymbol{Z}}}(\boldsymbol{z})$ is one when $\boldsymbol{z} \in \mathcal{G}_{\boldsymbol{Z}}$ and zero otherwise. It should be noted that since the boundary B in (6.5) is given by $\mathrm{VaR}_{99\%}(S)$ with $S = \sum_{i=1}^{d} Z_i$ then $\mathbb{P}[\boldsymbol{Z} \in \mathcal{G}_{\boldsymbol{Z}}] = 0.01$.

6.2.1.1 Reaching a rare event using intermediate steps

Instead of directly targeting the conditional distribution $(Z_1, \ldots, Z_d) \,|\, \{S \geq \mathrm{VaR}_{99\%}(S)\}$ the idea of the algorithm proposed in Chapter 4 and adapted to our current problem is to sequentially sample from intermediate distributions with conditioning events that become rarer until the point we reach the distribution of interest. The benefit of such an approach is that samples from a previous step (with a less rare conditioning event) are "guided" to the next algorithmic step (when targeting a rarer conditioning set) and, if carefully designed, no samples are wasted on the way to the target distribution, in the sense that no samples are incrementally weighted with a strictly zero weight.

In order to sample from the target distribution defined in (6.4) we use a sequence of intermediate distributions $\{\pi_t\}_{t=0}^{T}$, such that $\pi_T \equiv \pi$ and

$$\pi_t(\boldsymbol{z}) = f_{\boldsymbol{Z}}(\boldsymbol{z} \,|\, \boldsymbol{z} \in \mathcal{G}_{\boldsymbol{Z}_t}), \tag{6.6}$$

with $\mathcal{G}_{\boldsymbol{Z}_t} = \mathcal{G}_{\boldsymbol{Z}_t}(B_t)$ given by

$$\mathcal{G}_{\boldsymbol{Z}_t} = \left\{ \boldsymbol{z} \in \mathbb{R}^d \,:\, \sum_{i=1}^{d} z_i \geq B_t \right\}.$$

Remark 6.2.1. *Differently from Chapter 4, in order to make the algorithm more easily comparable with the one used for the conditional model, we do not transform the original random variable $\boldsymbol{Z}$ through its marginal distribution functions. Therefore, instead of sampling from the conditional copula we sample from the conditional joint distribution of $\boldsymbol{Z}$.*

The thresholds $B_1, \ldots, B_{T-1}$ are chosen in order to have increasingly rarer conditioning events as a function of t, starting from the unconditional joint density. In other words, $\{B_t\}_{t=0}^{T}$ needs to satisfy $0 = B_0 < \ldots < B_{T-1} < B_T = B = \mathrm{VaR}_{99\%}(S)$. Note that the choice $B_0 = 0$ assumes $S > 0$, $\mathbb{P}$-a.s., otherwise $B_0 = -\infty$. Depending on the choice of thresholds $\{B_t\}_{t=0}^{T-1}$ it

may be the case that the densities defined in (6.6) are only known up to a normalizing constant so, for now on, we work with γ_t, the unnormalized version of π_t:

$$\pi_t(\boldsymbol{z}) \propto \gamma_t(\boldsymbol{z}) = f_{\boldsymbol{Z}}(\boldsymbol{z}) \mathbb{1}_{\mathcal{G}_{\boldsymbol{Z}_t}}(\boldsymbol{z}). \tag{6.7}$$

If, at algorithmic time t, we have a set of N weighted samples $\{W_t^{(j)}, \boldsymbol{z}_t^{(j)}\}_{j=1}^N$ from π_t, with $\boldsymbol{z}_t^{(j)} = (z_{1,t}^{(j)}, \ldots, z_{d,t}^{(j)})$ then we construct the following empirical approximation:

$$\mathbb{E}[Z_i \,|\, S \geq B_t] \approx \sum_{j=1}^{N} W_t^{(j)} \boldsymbol{z}_{i,t}^{(j)}. \tag{6.8}$$

6.2.2 Allocations for the conditional model

From the discussion in Section 6.1 we see that the main difference between the marginalized and conditional models is the fact that the former density is analytically known whilst the latter is defined through an integral. In this section we discuss how to adapt the algorithm presented in Section 6.2.1 for situations where the target density cannot be analytically computed but a positive and unbiased estimator for it can be calculated.

Following the recent developments on *pseudo-marginal* methods (see [Finke, 2015] for a survey in the topic and Section 3.3.4, above) we substitute the unknown density $f_{\overline{\boldsymbol{Z}}}$ by a positive and unbiased estimate $\widehat{f}_{\overline{\boldsymbol{Z}}}$ and show the SMC procedure still targets the correct distribution — a strategy similar to the ones proposed in [Everitt et al., 2016] and [McGree et al., 2015]. In the context of rare event simulations a similar idea has been independently developed in [Vergé et al., 2016] where the authors study the impact of the parameter uncertainty in the probability of the rare event, whilst we analyse the impact in expectations conditional to the rare event (as in (6.3)).

To introduce the concept we first estimate $f_{\overline{\boldsymbol{Z}}}$ by $f_{\overline{\boldsymbol{Z}}}(\cdot \,|\, \boldsymbol{\theta})$, which can be seen as a "one sample" approximation to the integral in (5.7); then we show how to use an estimator based on $M \geq 1$ samples from $f_{\boldsymbol{\theta}}$. These two approaches have been named in the literature (see [Everitt et al., 2016] and references therein) as, respectively, the single auxiliary variable (SAV) and the multiple auxiliary variable (MAV) methods.

6.2.2.1 Single auxiliary variable method

To avoid direct use of $f_{\overline{\boldsymbol{Z}}}$ on the SMC sampler algorithm we provide a procedure on the joint space of $\overline{\boldsymbol{Z}}$ and the parameter $\boldsymbol{\theta}$, defined as $\mathcal{Y} = \mathbb{R}^d \times \boldsymbol{\Theta}$. The reader is referred to [Finke, 2015] for an exhaustive list of well known algorithms which can also be interpreted in a *extended space* way. The target distribution on this new space is defined as the joint distribution of $\overline{\boldsymbol{Z}}$ and $\boldsymbol{\theta}$ and its marginal with respect to $\overline{\boldsymbol{Z}}$ is precisely the density of the conditional model.

Formally, for $\boldsymbol{y} = (\overline{\boldsymbol{z}}, \boldsymbol{\theta})$, $\mathcal{G}_{\overline{\boldsymbol{Z}}}(\overline{B}) = \mathcal{G}_{\overline{\boldsymbol{Z}}} = \left\{ \overline{\boldsymbol{z}} \in \mathbb{R}^d \,:\, \sum_{i=1}^d \overline{z}_i \geq B \right\}$ and $\overline{B} = \mathrm{VaR}_{99\%}(\overline{S})$ we define

$$\pi^{\boldsymbol{y}}(\boldsymbol{y}) \propto \gamma^{\boldsymbol{y}}(\boldsymbol{y}) = f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}}}(\overline{\boldsymbol{z}}),$$

which has

$$\overline{\pi}(\overline{\boldsymbol{z}}) \propto \overline{\gamma}(\overline{\boldsymbol{z}}) = f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}}}(\overline{\boldsymbol{z}}) \tag{6.9}$$

as a marginal. Similarly to the densities defined in (6.7) and its version in the *path space* (see Equation (3.9)) we define a sequence of target distributions both in $\mathcal{Y}$ and $\mathcal{Y}^t$, respectively, as

$$\pi_t^{\boldsymbol{y}}(\boldsymbol{y}_t) \propto \gamma_t^{\boldsymbol{y}}(\boldsymbol{y}_t) = f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t)$$

and

$$\begin{aligned}\widetilde{\pi}_t^{\boldsymbol{y}}(\boldsymbol{y}_{1:t}) \propto \widetilde{\gamma}_t^{\boldsymbol{y}}(\boldsymbol{y}_{1:t}) &= \gamma_t^{\boldsymbol{y}}(\boldsymbol{y}_t) \prod_{s=1}^{t-1} L_s^{\boldsymbol{y}}(\boldsymbol{y}_{s+1}, \boldsymbol{y}_s) \\ &= f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t) \prod_{s=1}^{t-1} \overline{L}_s(\overline{\boldsymbol{z}}_{s+1}, \overline{\boldsymbol{z}}_s \,|\, \boldsymbol{\theta}_s) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_s),\end{aligned}$$

where the second identity specifies the choices of $L_s^{\boldsymbol{y}}$, in terms of $\overline{L}_s$ and $f_{\boldsymbol{\theta}}$.

Since we can perfectly sample from the distribution of $\boldsymbol{\theta}$, to move $\boldsymbol{y}$ samples backwards from time $s+1$ to s we split this process into sampling $\boldsymbol{\theta}_s$ from $f_{\boldsymbol{\theta}}$ (ignoring $\boldsymbol{\theta}_{s+1}$) and then, conditional on $\boldsymbol{\theta}_s$, moving $\overline{\boldsymbol{z}}_{s+1}$ to $\overline{\boldsymbol{z}}_s$. In other words, to sample

$$\boldsymbol{y}_s = (\overline{\boldsymbol{z}}_s, \boldsymbol{\theta}_s) \,\Big|\, \boldsymbol{y}_{s+1} = (\overline{\boldsymbol{z}}_{s+1}, \boldsymbol{\theta}_{s+1}) \sim L_s^{\boldsymbol{y}}(\boldsymbol{y}_{s+1}, \boldsymbol{y}_s),$$

we split the process in two stages,

1. $\boldsymbol{\theta}_s \sim f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_s)$;
2. $\overline{\boldsymbol{z}}_s \,|\, \overline{\boldsymbol{z}}_{s+1} \sim \overline{L}_s(\overline{\boldsymbol{z}}_{s+1}, \overline{\boldsymbol{z}}_s \,|\, \boldsymbol{\theta}_s)$

The importance distribution on the path space of $\boldsymbol{y}$ can, then, be expressed according to

$$\begin{aligned}\widetilde{q}_t^{\boldsymbol{y}}(\boldsymbol{y}_{1:t}) &= q_1^{\boldsymbol{y}}(\boldsymbol{y}_1) \prod_{s=2}^{t} K_s^{\boldsymbol{y}}(\boldsymbol{y}_{s-1}, \boldsymbol{y}_s) \\ &= \overline{q}_1(\overline{\boldsymbol{z}}_1) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1) \prod_{s=2}^{t} \overline{K}_s(\overline{\boldsymbol{z}}_{s-1}, \overline{\boldsymbol{z}}_s \,|\, \boldsymbol{\theta}_s) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_s)\end{aligned}$$

and, once again, the second identity provides the choices of $q_1^{\boldsymbol{y}}$ and $K_s^{\boldsymbol{y}}$, i.e.,

$$q_1^{\boldsymbol{y}}(\boldsymbol{y}_1) = \overline{q}_1(\overline{\boldsymbol{z}}_1) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_1) \quad \text{and} \quad K_s^{\boldsymbol{y}}(\boldsymbol{y}_{s-1}, \boldsymbol{y}_s) = \overline{K}_s(\overline{\boldsymbol{z}}_{s-1}, \overline{\boldsymbol{z}}_s \,|\, \boldsymbol{\theta}_s) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_s).$$

Therefore, a SMC procedure targeting the sequence $\{\pi_t^{\boldsymbol{y}}(\boldsymbol{y}_t)\}_{t=1}^T$ produces unnormalized weights

$$\begin{aligned}w_t^{\boldsymbol{y}} &= \frac{\widetilde{\gamma}_t^{\boldsymbol{y}}(\boldsymbol{y}_{1:t})}{\widetilde{q}_t^{\boldsymbol{y}}(\boldsymbol{y}_{1:t})} \\ &= w_{t-1}^{\boldsymbol{y}} \frac{\gamma_t^{\boldsymbol{y}}(\boldsymbol{y}_t) L_{t-1}^{\boldsymbol{y}}(\boldsymbol{y}_t, \boldsymbol{y}_{t-1})}{\gamma_{t-1}^{\boldsymbol{y}}(\boldsymbol{y}_{t-1}) K_t^{\boldsymbol{y}}(\boldsymbol{y}_{t-1}, \boldsymbol{y}_t)} \\ &= w_{t-1}^{\boldsymbol{y}} \frac{f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t) \overline{L}_{t-1}(\overline{\boldsymbol{z}}_t, \overline{\boldsymbol{z}}_{t-1} \,|\, \boldsymbol{\theta}_{t-1}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{t-1})}{f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_{t-1} \,|\, \boldsymbol{\theta}_{t-1}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{t-1}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_{t-1}}}(\overline{\boldsymbol{z}}_{t-1}) \overline{K}_t(\overline{\boldsymbol{z}}_{t-1}, \overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}_t)} \\ &= w_{t-1}^{\boldsymbol{y}} \frac{f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t) \overline{L}_{t-1}(\overline{\boldsymbol{z}}_t, \overline{\boldsymbol{z}}_{t-1} \,|\, \boldsymbol{\theta}_{t-1})}{f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_{t-1} \,|\, \boldsymbol{\theta}_{t-1}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_{t-1}}}(\overline{\boldsymbol{z}}_{t-1}) \overline{K}_t(\overline{\boldsymbol{z}}_{t-1}, \overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t)},\end{aligned}$$

that can be used to create weighted samples from $\overline{\pi}_t(\overline{\boldsymbol{z}}_t)$, which is the desired marginal of $\pi_t^{\boldsymbol{y}}(\boldsymbol{y}_t)$, the density required for the capital allocation.

Remark 6.2.2. *From the structure of the mutation kernels $K_t^{\boldsymbol{y}}$ it should be noticed that at each iteration t a new value of $\boldsymbol{\theta}_t$ needs to be generated and used to sample $\overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\theta}_t$. In other words, for each particle $j = 1, \ldots, N$ a different $\boldsymbol{\theta}_t^{(j)}$ is to be used for each $\overline{\boldsymbol{z}}_t^{(j)} \,|\, \boldsymbol{\theta}_t^{(j)}$.*

6.2.2.2 Multiple auxiliary variable

In the previous algorithm we, indirectly, estimate the density $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ by $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta})$. In this section we discuss how to use a different and more robust estimator, using $M \geq 1$ samples from $\boldsymbol{\theta}$. In the context of pseudo-marginal Monte Carlo Markov Chain (MCMC) [Andrieu and Vihola, 2015] show that reducing the variance of the estimate of the unknown density $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ leads to reduced asymptotic variance of estimators from the MCMC. For SMC algorithms this strategy has been used, for example, in [McGree et al., 2015] and [Everitt et al., 2016].

Before proceeding, we note that even in the case that $M = 1$ the algorithm still produces asymptotic and unbiased estimators (when $N \to +\infty$). However, the rate of variance reduction in this asymptotic is directly affected by the choice of M, in a non-trivial manner. Furthermore, the asymptotic variance of Central Limit Theorem (CLT) estimators under the class of such pseudo-marginal Monte Carlo approaches is strictly ordered in M, with M increasing reducing the the asymptotic variance.

For any $M \geq 1$, a positive and unbiased estimate for $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ can be constructed as

$$\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta}) = \frac{1}{M} \sum_{i=1}^{M} f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}^{(i)}), \tag{6.10}$$

where $\boldsymbol{\vartheta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}) \in \Theta^M$ and each $\boldsymbol{\theta}^{(m)}$ is sampled independently from $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. Note that when only one sample of $\boldsymbol{\theta}$ is used to estimate $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ the estimator is reduced to $\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta}) = f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta})$. Also, note that $\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta}) \to f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ point-wise when $M \to +\infty$, by the law of large numbers.

Since the random variable $\boldsymbol{\vartheta}$ has density $f_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) = \prod_{i=1}^{M} f_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{(i)})$, then

$$\begin{aligned}
\int_{\boldsymbol{\Theta}^M} \widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta}) f_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta} &= \int_{\boldsymbol{\Theta}^M} \frac{1}{M} \sum_{i=1}^{M} f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}^{(i)}) \prod_{i=1}^{M} f_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{(i)}) d\boldsymbol{\theta}^{(1)} \ldots d\boldsymbol{\theta}^{(M)} \\
&= \int_{\boldsymbol{\Theta}} f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta} = f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}).
\end{aligned}$$

Therefore the density $\overline{\pi}(\overline{\boldsymbol{z}})$ constructed in (6.9) is the marginal of the new target density defined on $\mathcal{Y}_M = \mathbb{R}^d \times \boldsymbol{\Theta}^M$

$$\pi^{\boldsymbol{y}}(\boldsymbol{y};\, \boldsymbol{\vartheta}) \propto \gamma^{\boldsymbol{y}}(\boldsymbol{y};\, \boldsymbol{\vartheta}) = \widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta}) f_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}}}(\overline{\boldsymbol{z}}).$$

Apart from the cumbersome notation, the same argument from the previous section can be used to show that a SMC procedure with estimated density $\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}};\, \boldsymbol{\vartheta})$ replacing $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ has unnormalized weights given by

$$w_t^{\boldsymbol{y}} = w_{t-1}^{\boldsymbol{y}} \frac{\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t;\, \boldsymbol{\vartheta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t) \overline{L}_{t-1}(\overline{\boldsymbol{z}}_t,\, \overline{\boldsymbol{z}}_{t-1} \,|\, \boldsymbol{\vartheta}_{t-1})}{\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_{t-1};\, \boldsymbol{\vartheta}_{t-1}) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_{t-1}}}(\overline{\boldsymbol{z}}_{t-1}) \overline{K}_t(\overline{\boldsymbol{z}}_{t-1},\, \overline{\boldsymbol{z}}_t \,|\, \boldsymbol{\vartheta}_t)},$$

when targeting a sequence $\{\overline{\pi}_t(\overline{\boldsymbol{z}}_t)\}_{t=1}^T$ with $\overline{\pi}_T(\overline{\boldsymbol{z}}_T) = \overline{\pi}(\overline{\boldsymbol{z}})$.

LoB	Reserves	Premium
1 MTPL	2,391.64	503.14
2 Motor Hull	99.08	573.26
3 Property	449.26	748.76
4 Liability	870.27	299.73
5 Workers Compensation (UVG)	1,104.66	338.63
6 Commercial Health	271.54	254.21
7 Private Health	7.32	7.2
8 Credit and Surety	49.5	34.64
9 Others	67.64	46.28
Total	5,310.92	2,805.87

Table 6.1: Original balance sheet.

6.3 Data description and parameter estimation

Starting from the balance sheet of a fictitious insurance company, in this section we discuss how we set up the fixed parameters in the models discussed so far. Using this balance sheet and the information contained in the SST we also show how to generate realistic claims triangles and, based on them, how to perform Bayesian inference for the unknown parameters. Our starting point is the fictitious balance sheet shown in Table 6.1, which is intended to represent a large insurance company in Switzerland (for this reason, in this Chapter all monetary units should be understood as millions of Swiss Francs (CHF)).

6.3.1 Hyperparameters for ϕ_j

Based on SST's standard runoff pattern (see Table 6.2), which is the cumulated proportion of paid claims for the specific accident year, we first compute the implied CL factors $f_j^{(\ell)}$ as follows (once again we suppress the index ℓ of the LoB). If F_j is the *cumulative* claims payment pattern for development year j we define

$$f_j = \frac{F_{j+1}}{F_j}, \text{ for } j = 0, \dots, J-1.$$

These values can, then, be used as a hyperparameter in the prior for ϕ_j (see Model Assumptions 5.3.1, item (c)).

To generate data from the model (see Section 6.3.3) we fix $\phi_j = 1/f_j$ and $\sigma_j = s_j/f_j$, where s_j is Mack's standard deviation estimate calculated from exogenous triangles. The values of s_j are presented in Table 6.3.

6.3.2 Current year small and large claims

To calculate the expected number of CY claims, λ_{CY}, defined in Section 5.4, we first set what we believe to be the *claims ratio* for each LoB, i.e., how much of the premium in that LoB is used to cover incoming claims (all the rest covers business' costs). This information is available in Table 6.4, along with the *average claim amount*. Based on these values the expected number

LoB	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11	Year 12	Year 13	Year 14	Year 15
1	30.18%	15.63%	5.78%	4.94%	4.43%	4.34%	4.09%	3.92%	3.66%	3.5%	3.08%	2.64%	2.16%	1.86%	1.5%	1.3%
2	81.08%	18.67%	0.24%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3	58.24%	35.06%	4.36%	1.37%	0.64%	0.33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	26.55%	23.53%	8.33%	6.18%	4.79%	4.15%	3.63%	3.14%	2.55%	2.11%	1.8%	1.59%	1.35%	1.2%	1.12%	1.02%
5	40.62%	24.92%	7.14%	4.86%	4.43%	3.13%	2.57%	1.67%	1.31%	1.22%	1.05%	0.69%	0.6%	0.56%	0.51%	0.47%
6	36.83%	47.68%	14.2%	0.88%	0.28%	0.14%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	46.26%	38.05%	10.78%	2.94%	1.27%	0.69%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	45.85%	35.28%	11.35%	3.72%	1.62%	0.91%	0.52%	0.32%	0.2%	0.13%	0.1%	0%	0%	0%	0%	0%
9	58.24%	35.06%	4.36%	1.37%	0.64%	0.33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

LoB	Year 16	Year 17	Year 18	Year 19	Year 20	Year 21	Year 22	Year 23	Year 24	Year 25	Year 26	Year 27	Year 28	Year 29	Year 30
1	1.06%	0.88%	0.73%	0.64%	0.6%	0.53%	0.47%	0.44%	0.41%	0.37%	0.29%	0.21%	0.15%	0.12%	0.1%
2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	0.88%	0.77%	0.72%	0.66%	0.6%	0.55%	0.52%	0.49%	0.45%	0.4%	0.31%	0.22%	0.16%	0.13%	0.11%
5	0.43%	0.4%	0.37%	0.35%	0.33%	0.31%	0.29%	0.27%	0.26%	0.24%	0.23%	0.22%	0.2%	0.19%	0.18%
6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 6.2: SST's (2015) standard development patterns for claims provision (normalized to have only 30 development years and then rounded to 2 digits).

LoB	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Year 11	Year 12	Year 13	Year 14
1	0.5673	0.228	0.1922	0.2681	0.2683	0.3949	0.2652	0.2641	0.2789	0.3055	0.1458	0.1577	0.214	0.1001	0.1016
2	0.664	0.0659													
3	1.3614	0.4921	0.3215	0.0875	0.0666										
4	0.8248	0.4328	0.4021	0.3644	0.3772	0.2729	0.5268	0.244	0.2786	0.1559	0.266	0.0776	0.0757	0.122	0.0418
5	0.9914	0.3317	0.1807	0.1072	0.074	0.0444	0.0359	0.0255	0.019	0.0106	0.0166	0.0094	0.004	0.0105	0.004
6	0.6069	0.2405	0.0597	0.0371	0.0172										
7	0.1053	0.045	0.0157	0.0113	0.0091										
8	0.3098	0.0737	0.031	0.0203	0.0137	0.0051	0.002	0.0026	0.002	0.0014	0.0011				
9	0.9163	0.191	0.1248	0.034	0.0258										

LoB	Year 15	Year 16	Year 17	Year 18	Year 19	Year 20	Year 21	Year 22	Year 23	Year 24	Year 25	Year 26	Year 27	Year 28	Year 29
1	0.0466	0.1097	0.1081	0.0583	0.1353	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916
2															
3															
4	0.0272	0.0886	0.0422	0.019	0.0238	0.019	0.0152	0.0122	0.0097	0.0078	0.0062	0.005	0.004	0.0032	0.0025
5	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
6															
7															
8															
9															

Table 6.3: Mack's standard deviation estimates, s_j, based on exogenous triangles.

LoB	Claims ratio	Average claim amount	Market share
1	90%	0.005	
2	75%	0.003	20%
3	75%	0.004	
4	75%	0.004	
5	90%	0.004	10%
6	90%	0.003	
7	90%	0.002	
8	80%	0.003	
9	80%	0.003	

Table 6.4: Claims ratio, average claim amount (in millions of CHF) and market share.

of claims is defined as

$$\lambda_{CY} = \frac{\text{Claims ratio} \times \text{Premium}}{\text{Average claim amount}}.$$

In light of the expected number of CY claims its value is used to compute the expected number of individual large claims, $\lambda_{CY,l}$, as in (5.38). Using the fact that $\lambda_{CY,s} = \lambda_{CY} - \lambda_{CY,l}$ we calculate the coefficient of variation for small CY claims given in (5.35).

The last ingredient in Model Assumptions 5.4.1 is $\mathbb{E}[Z_{CY,s} \,|\, \mathcal{D}(t)]$ which is given by

$$\mathbb{E}[Z_{CY,s} \,|\, \mathcal{D}(t)] = \text{Claims ratio} \times \text{Premium} - \mathbb{E}[Z_{CY,l} \,|\, \mathcal{D}(t)],$$

and the expectation on the right hand side is given either in Model Assumptions 5.4.2 or Model Assumptions 5.4.4, depending on the LoB.

For the Large Claims from Model Assumptions 5.4.2 and 5.4.4 we assume the threshold for large claims β to be equal to 5 (millions of CHF). For the large cumulated claims we use LoBs market share as given in Table 6.4.

The resulting parameters can be found in Table 6.5. Note these parameters are the same both for the marginalized and conditional models.

6.3.3 Data generating process

In this section we present the process used to generate claims triangles using the balance sheet data from Table 6.1 in a way that the estimated reserves from the data match, as closely as possible, the reserves from Table 6.1.

First of all, for each LoB we set the maximum number of development years as the number of years it takes until $F_j = 1$, where F_j denotes the *cumulative* payment pattern for development year j (see Section 6.3.1). In order not to use too old claims, we assume each LoB has at most 10 accident years and set $I = \max(J + 1, 10)$.

For different accident years we calculate the present value of the runoff pattern, using a constant claim inflation $r = 2\%$ for all years and LoBs. More precisely, we have that

$$PV_i(F_j) = (1 + r)^{-i} F_j \quad \text{for} \quad j = 1, \ldots, J \quad \text{and} \quad j + i > I.$$

	Reserve /					Standalone		Marginalized			Conditional		
LoB	Premium	σ	μ	CoVa	Expectation	ES	SCR	ES	SCR	Div. benefit	ES	SCR	Div. benefit
1	2365.44	0.0287	7.7659	2.87%	2365.44	2546.31	180.87	2489.85	124.41	31.22%	2492.05	126.61	30%
2	99.37	0.2164	4.5755	21.9%	99.37	173.23	73.86	131.73	32.36	56.19%	132.59	33.21	55.03%
3	405.99	0.1142	5.9998	11.46%	405.99	547.25	141.26	479.11	73.12	48.24%	485.27	79.28	43.88%
4	870.19	0.0315	6.7682	3.15%	870.19	946.06	75.87	905.48	35.29	53.49%	905.29	35.1	53.73%
5	1105.95	0.0193	7.0083	1.93%	1105.95	1164.04	58.09	1137.06	31.11	46.44%	1136.88	30.93	46.76%
6	274.91	0.041	5.6156	4.1%	274.91	306.43	31.52	287.33	12.42	60.59%	286.97	12.06	61.74%
7	7.15	0.0547	1.9657	5.48%	7.15	8.26	1.11	7.45	0.3	73.27%	7.43	0.28	74.5%
8	48.18	0.0493	3.8738	4.93%	48.18	54.89	6.71	50.51	2.32	65.36%	50.43	2.25	66.44%
9	72.2	0.1332	4.2706	13.38%	72.2	102.16	29.96	85.32	13.12	56.21%	85.15	12.95	56.77%
Total PY	5249.38				5249.38	5848.63	599.25	5573.84	324.45	45.86%	5582.06	332.67	44.49%
1	503.14	0.0685	6.0958	6.86%	448.94	533.07	84.13	499.16	50.21	40.32%	498.37	49.43	41.25%
2	573.26	0.0702	6.0356	7.03%	402.87	504.2	101.33	472.25	69.38	31.53%	471.66	68.79	32.11%
3	748.76	0.0683	6.3013	6.84%	547.23	654.38	107.15	603.36	56.13	47.62%	602.61	55.38	48.31%
4	299.73	0.0923	5.3596	9.25%	216.7	272.05	55.35	239.69	22.99	58.47%	239.57	22.87	58.69%
5	338.63	0.0648	5.6841	6.49%	303.77	349.69	45.92	319.17	15.4	66.47%	318.71	14.94	67.45%
6	254.21	0.0804	5.4296	8.05%	228.79	282.62	53.83	249.63	20.85	61.28%	249.31	20.52	61.88%
7	7.2	0.1047	1.8628	10.5%	6.48	8.52	2.04	7.01	0.53	73.84%	7.01	0.53	74.06%
8	34.64	0.0981	3.3172	9.84%	27.72	35.84	8.13	30.32	2.6	67.95%	30.28	2.57	68.44%
9	46.28	0.1004	3.6066	10.06%	37.03	48.16	11.14	41.83	4.81	56.83%	41.79	4.77	57.19%
Total CY,s	2805.85				2219.53	2688.53	469.02	2462.42	242.9	48.21%	2459.31	239.8	48.87%

Table 6.5: Parameters and capital calculations for the marginalized and conditional models (Part I/II).

					Standalone		Marginalized			Conditional		
LoB	$\beta^{(5)}$	γ	α	Expectation	ES	SCR	ES	SCR	Div. benefit	ES	SCR	Div. benefit
1	2.5		2.8	3.89	20.14	16.25	4.03	0.15	99.1%	4.01	0.12	99.27%
2	13.35	300	1.85	27.08	191.21	164.13	39.96	12.88	92.15%	39.61	12.53	92.36%
3	6.28	100	1.5	14.34	84.31	69.97	16.5	2.16	96.91%	16.45	2.11	96.98%
4	3.88	100	1.8	8.1	61.34	53.24	8.94	0.84	98.42%	8.91	0.81	98.48%
5	0.5		2	1	10	9	1.07	0.07	99.19%	1.12	0.12	98.69%
Total CY,l				54.41	367	312.59	70.5	16.1	94.85%	70.1	15.69	94.98%
Total	8055.26			7523.32	8904.18	1380.86	8106.77	583.45	57.75%	8111.5	588.18	57.4%

Table 6.5: (Continued) Parameters and capital calculations for the marginalized and conditional models (Part II/II).

For the most recent accident year, $i = I$, we define the expected ultimate claim by

$$C^*_{I,J} = R \times \frac{\sum_{j=1}^{J} P_{I,j}}{\sum_{j=1}^{J} F_j},$$

where R denotes the reserves from Table 6.1 and

$$P_{I,j} = \frac{PV_I(F_j)}{\sum_{i=1}^{I}\sum_{j=1}^{J} PV_i(F_j)}.$$

Note that $C^*_{I,J}$ is neither the ultimate claim predictor for the conditional model defined in (5.8) nor the marginalized one from (5.14). In this context $C^*_{I,J}$ is just an auxiliary variable being used in order to simulate triangles which have estimated reserves similar to the original ones in Table 6.1.

For the remaining accident years the expected ultimate claim is taken as the present value of $C^*_{I,J}$. In other words,

$$C^*_{i,J} = PV_{i-I}(C^*_{I,J}) = (1+r)^{I-i} C^*_{I,J}.$$

Given all the values of $C^*_{i,J}$, we compute $E^*_i = F_0 \times C^*_{i,J}$, the expected initial payment for each accident year. These values are, then, combined with the coefficients of variation for CY small claims and used to simulate the first column of our triangles as

$$C_{i,0} \sim LN(m^*_i, V^*_i),$$

with the auxiliary parameters $m^*_i = \log(E^*_i) - V^*_i/2$, $V^*_i = \log(1 + \text{CoVa}^2_{CY})$ and CoVa_{CY} the coefficient of variation of CY small claims, based on Model Assumptions 5.4.1. For the remaining development years we follow Model Assumptions 5.3.1 (a) with $\phi_j = 1/f_j$ and $\sigma_j = s_j/f_j$, as discussed in Section 6.3.1.

Figure 6.1 presents the generated cumulative claims payments for all LoBs, where each line represents the cumulative claims payment. In each plot the lighter colours represent more recent accident years which are not yet fully developed. The reserves calculated based on this dataset are presented in Table 6.5 and given these values the original reserves from Table 6.1 are ignored.

6.3.4 Parameter estimation

In order to estimate the variance parameters σ_j in Model Assumptions 5.3.1 we assume priors centred at Mack's [Mack, 1993] CL standard deviation estimator normalized by the CL factor f, both implied by the data. Formally,

$$\widehat{\sigma}_j(t) = \frac{\sqrt{\widehat{s}^2_j(t)}}{\widehat{f}_j(t)}, \ \forall 0 \le j \le J-1, \tag{6.11}$$

where $\widehat{s}^2_{J-1}(t) = \min\{\widehat{s}^2_{J-3}(t), \widehat{s}^2_{J-2}(t), \widehat{s}^4_{J-2}(t)/\widehat{s}^2_{J-3}(t)\} = \min\{\widehat{s}^2_{J-3}(t), \widehat{s}^4_{J-2}(t)/\widehat{s}^2_{J-3}(t)\}$.

To generate samples from the posteriors we use a Metropolis-Hastings algorithm, with proposals given by a truncated Normal centred at the current point and standard deviation equal

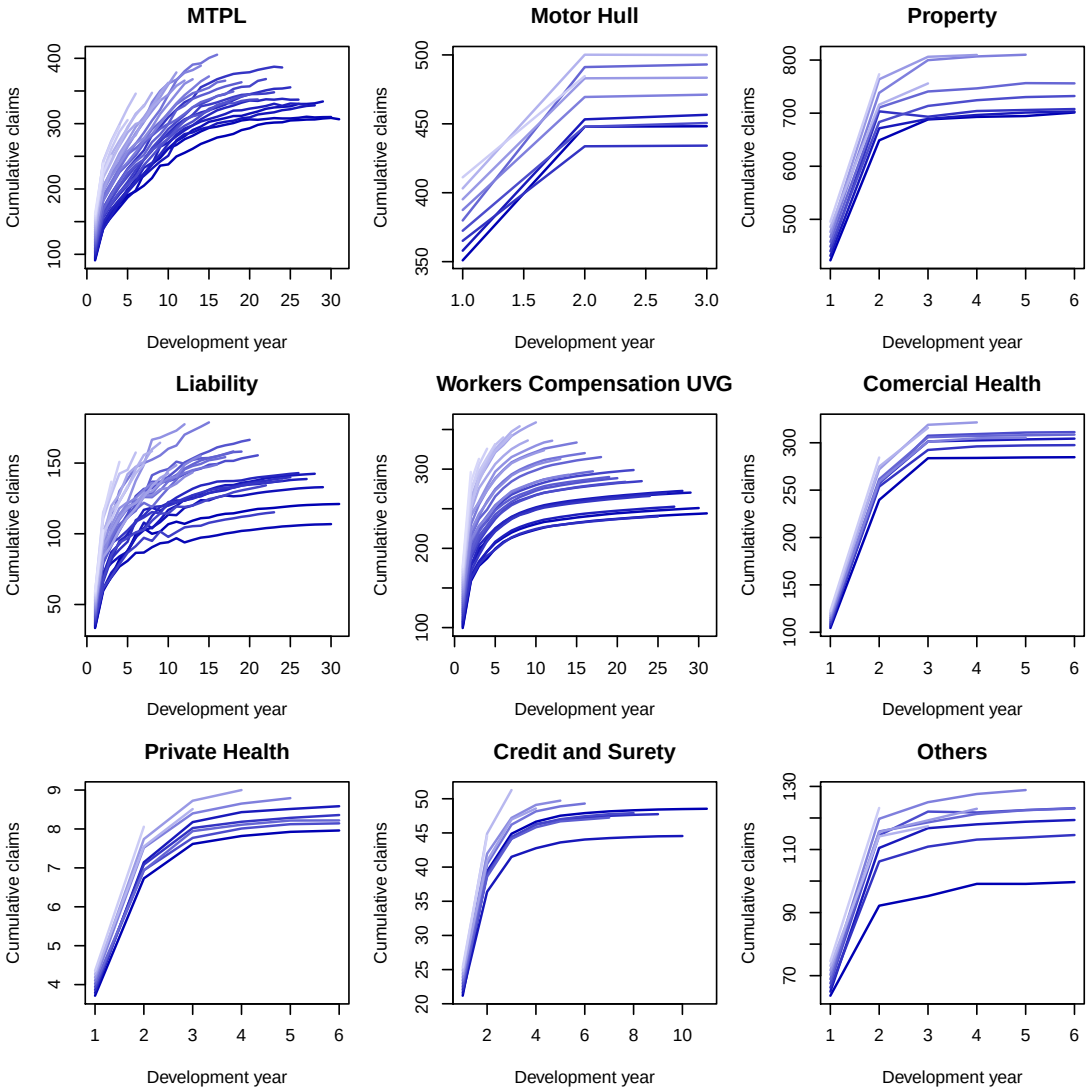

Figure 6.1: Cumulative claims payment (in millions of CHF). Lighter colours represent more recent accident years.

to $10\times d_j$. All the chains are started at the CL variance estimate and the upper limit for the prior, $d_j = k \times \widehat{\sigma}_j(t)$ is set as $k = 5$ times the CL variance estimate. To be left with $N_{MCMC} = 1,000$ samples from the posterior we ran the Markov chains for 12,500 iterations, discarding the first 20% as a burn-in and keeping every 10th iteration of the remaining simulations.

Some of the results are presented in Figures 6.2, 6.3 and 6.4 where one finds the unnormalized posteriors, the histogram of the MCMC outputs and a red dashed line indicating the CL variance estimate for three different LoBs: MTPL, Motor Hull and Property, respectively. As expected for unidimensional and unimodal densities the resulting estimates are highly accurate. It is also worth noticing that the larger the development year j the more diffuse the posterior is, due to the diminishing amount of data available. In the limit, when $j = J - 1$ the information available is not enough to estimate the variance parameter and, therefore, as can be seen from the posterior distribution derived in (5.25), the posterior is the same as the prior.

Using the sample of size $N_{MCMC} = 1,000$ mentioned above, the calculated parameters for the marginalized model are presented in Table 6.5. For the conditional model we use the same sample from the posterior and calculate the one value of σ_{PY} and μ_{PY} for each sampled value $\boldsymbol{\sigma}$. The resulting (transformed) samples are presented as histograms in Figures 6.5 and 6.6 and, for comparison only, the relevant marginalized parameters are included as a red dashed line.

6.3.5 The correlation matrices

For the copula correlation matrices we follow the procedures outlined in Sections 5.5.1 and 5.5.2. The resulting matrix for the marginalized model is found in Table 6.6. From [FINMA, 2016] it can be seen the values in $\boldsymbol{\Omega}_{PY,\,CY,s}$ are very similar to ones in the standard $\boldsymbol{\Lambda}_{PY,\,CY,s}$.

Also, it worth noticing that differently from SST's original correlation matrix, the block $\boldsymbol{\Omega}_{PY,\,CY,s}$ is no longer symmetric, i.e., in order to have $\text{Corr}(Z^{(1)}_{PY},\, Z^{(2)}_{CY} \,|\, \mathcal{D}(t)) = \text{Corr}(Z^{(2)}_{PY},\, Z^{(1)}_{CY} \,|\, \mathcal{D}(t))$ the term $(1,2)$ of the matrix $\boldsymbol{\Omega}_{PY,\,CY,s}$ is not equal to the term $(2,1)$ of the same matrix.

The results for the copula correlation $\overline{\boldsymbol{\Omega}}_{PY,\,CY,s}$ follow the same patterns as $\boldsymbol{\Omega}_{PY,\,CY,s}$ and for this reason its values are omitted.

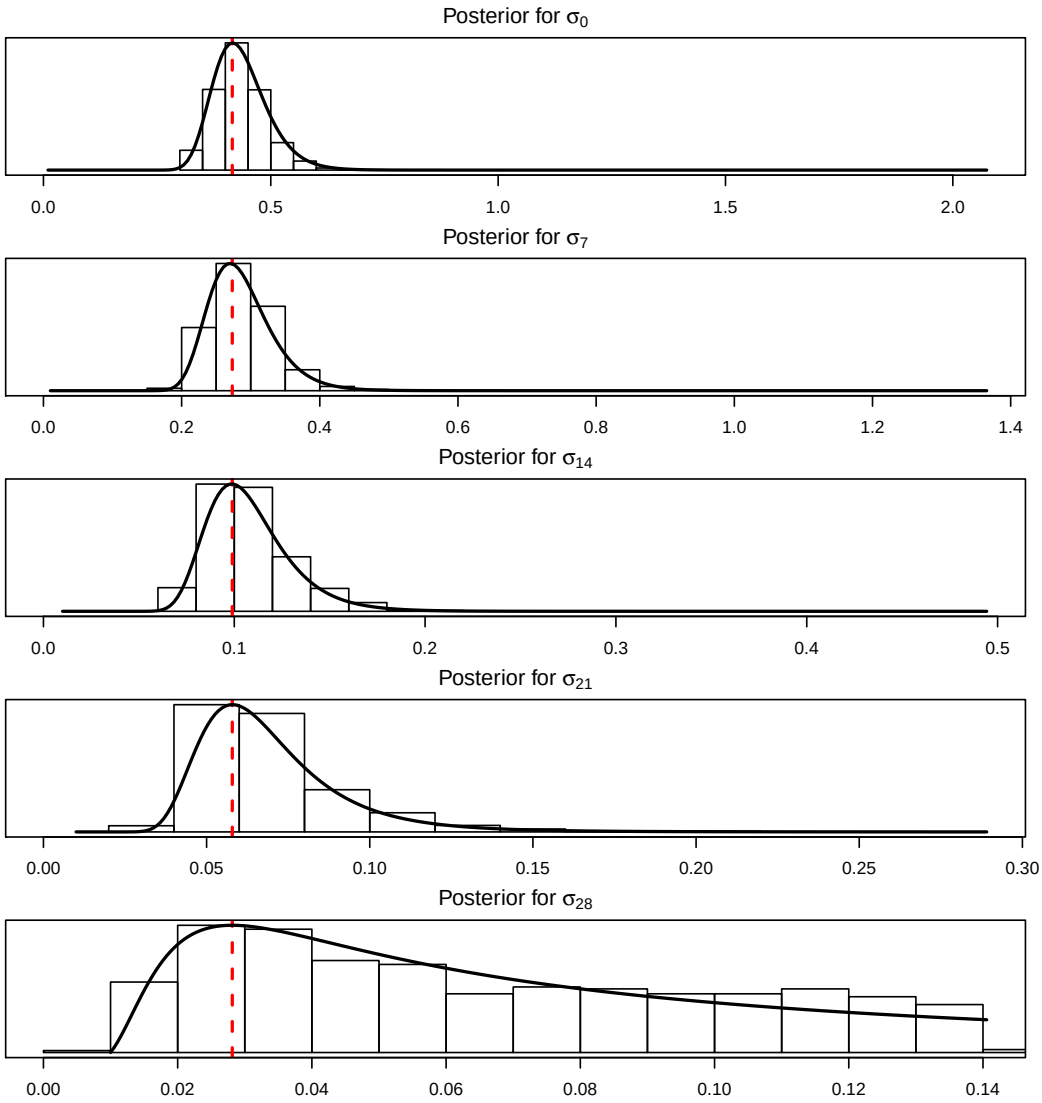

Figure 6.2: Posterior distributions for σ_j for the MTPL line of business. One sees solid lines representing the unnormalized posteriors, the histogram of the MCMC outputs and a red dashed line indicating the CL variance estimate.

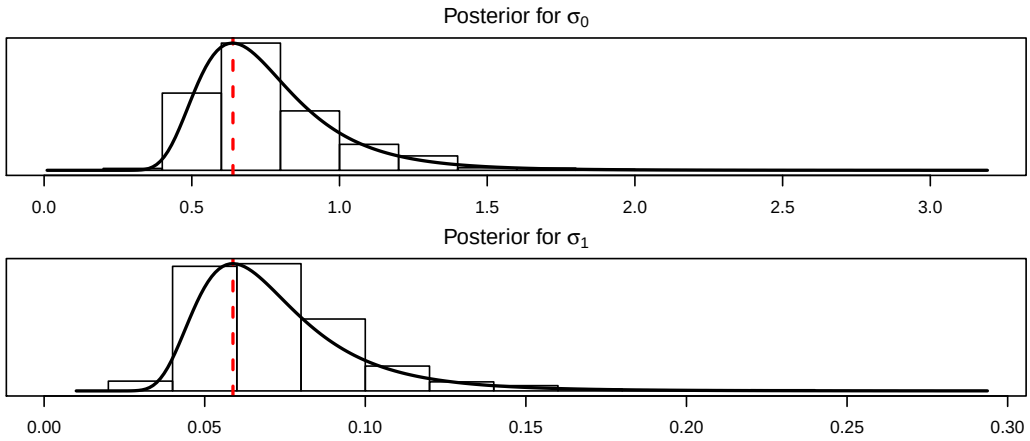

Figure 6.3: Posterior distributions for σ_j for the Motor Hull line of business. One sees solid lines representing the unnormalized posteriors, the histogram of the MCMC outputs and a red dashed line indicating the CL variance estimate.

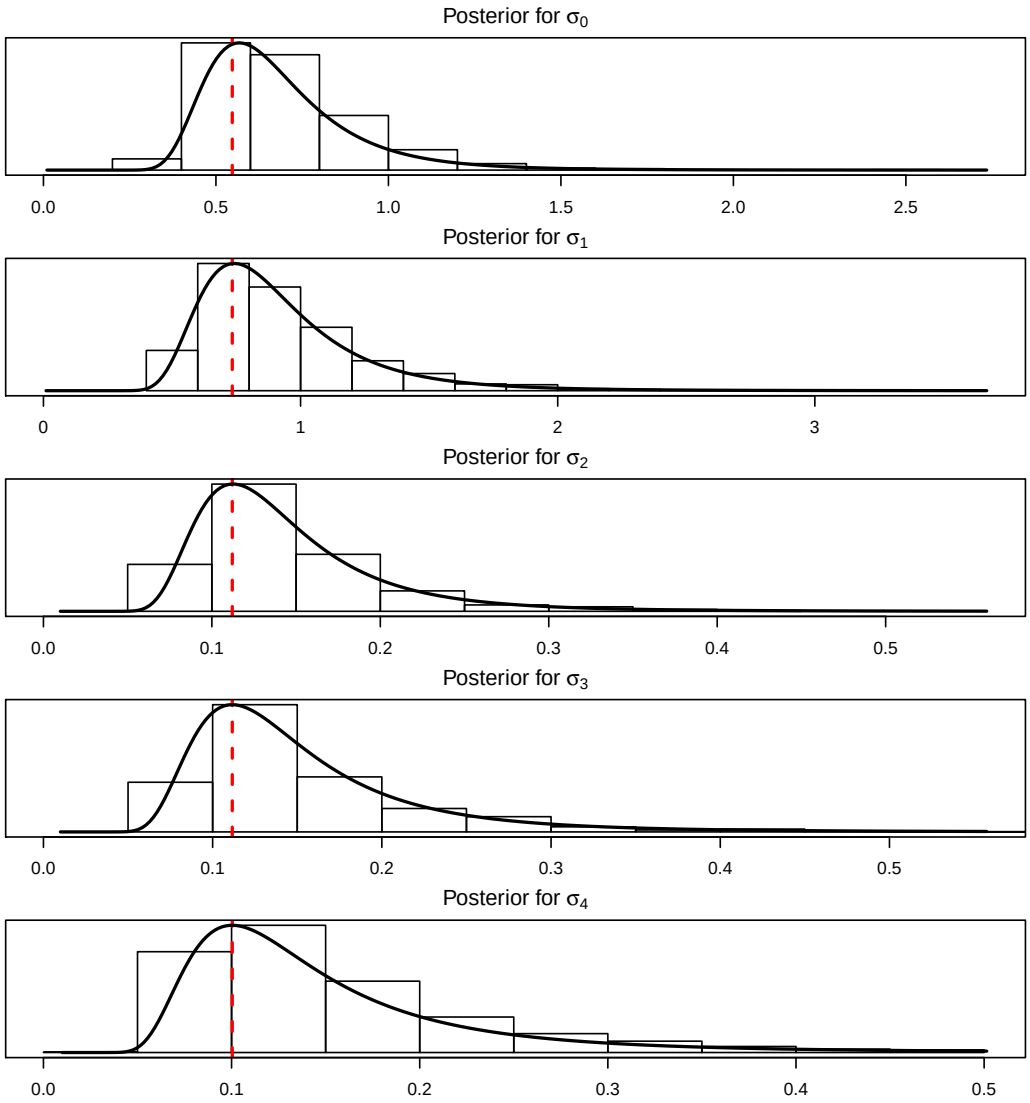

Figure 6.4: Posterior distributions for σ_j for the Property line of business. One sees solid lines representing the unnormalized posteriors, the histogram of the MCMC outputs and a red dashed line indicating the CL variance estimate.

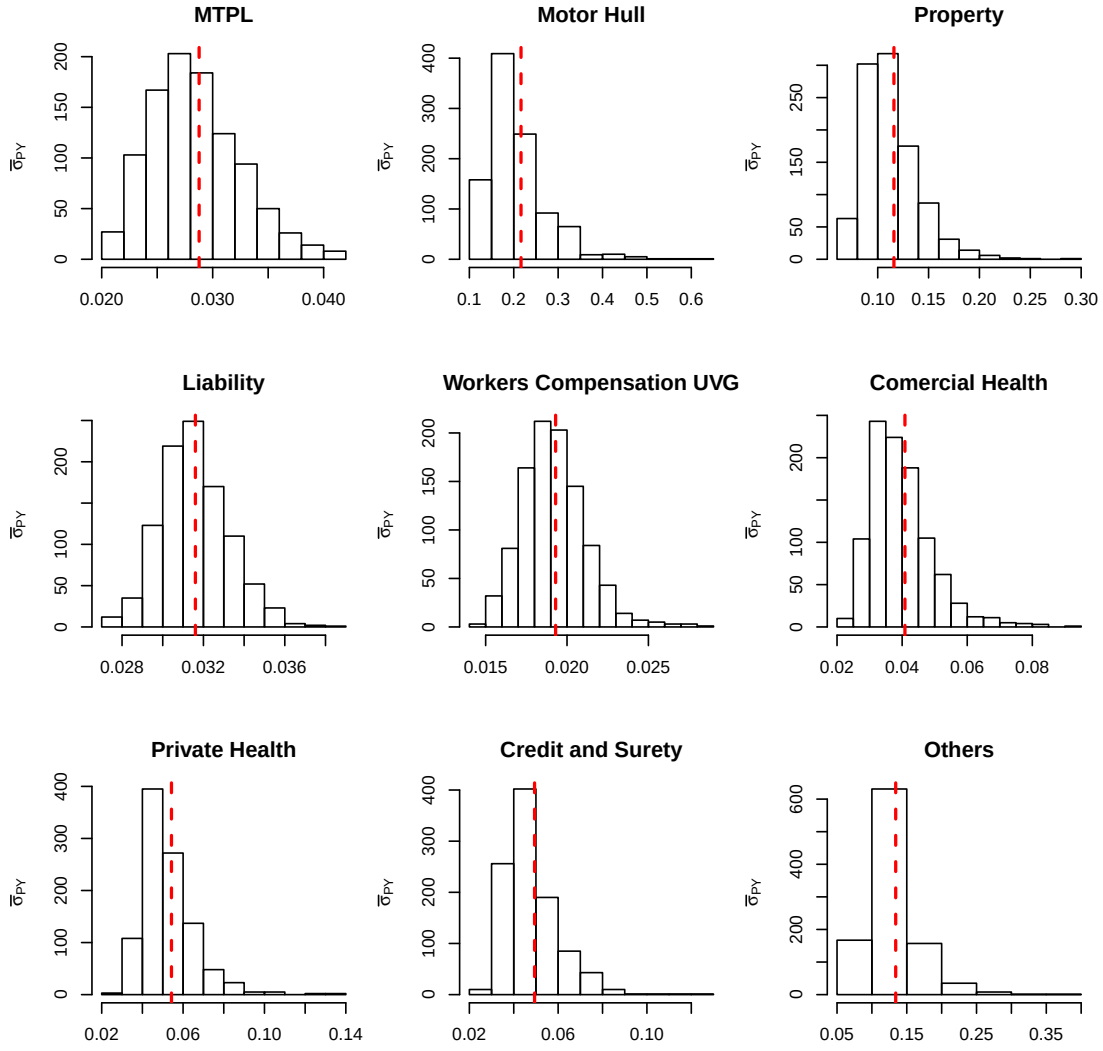

Figure 6.5: Histogram of the parameter $\overline{\sigma}_{PY}$ for the conditional model. Red dashed line: σ_{PY}.

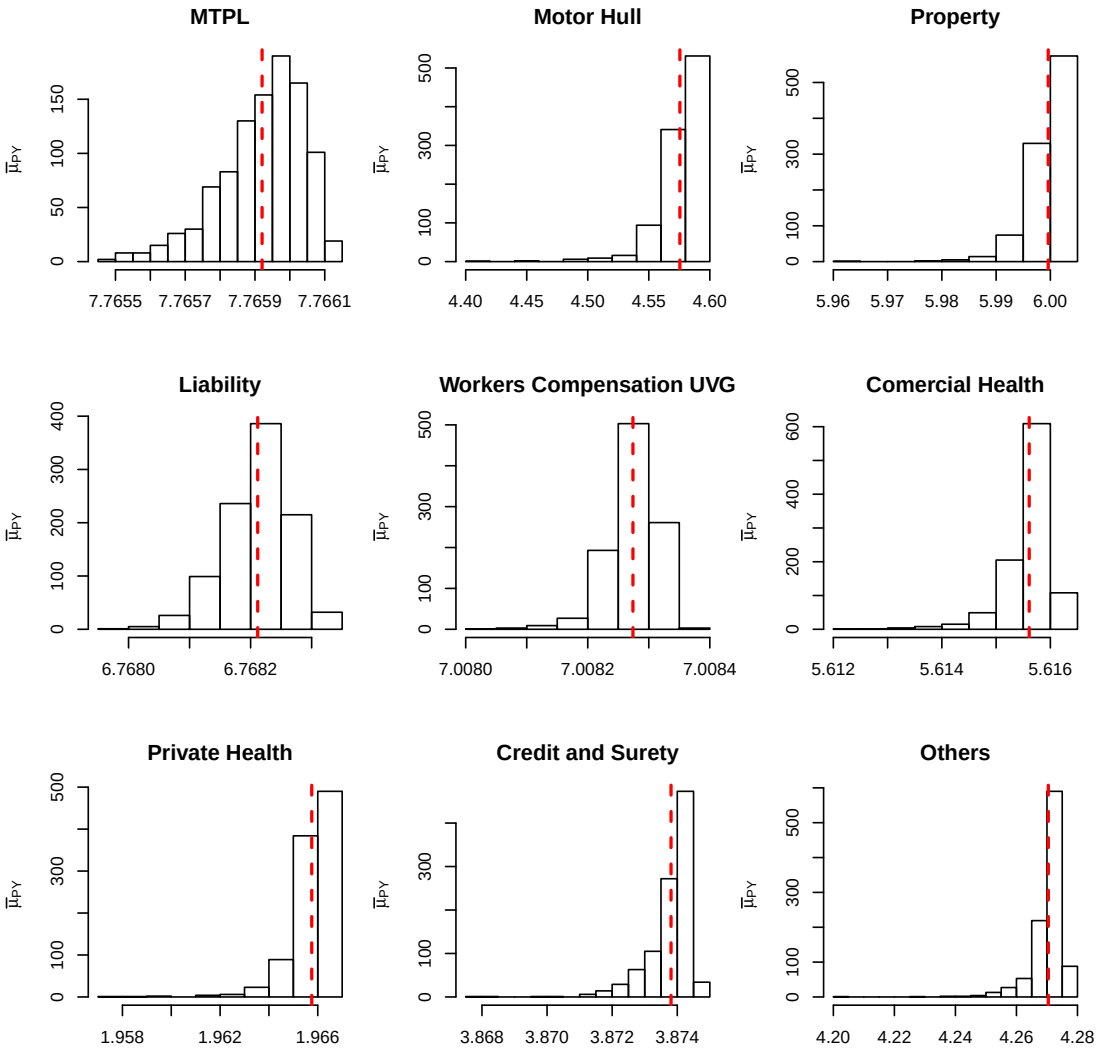

Figure 6.6: Histogram of the parameter $\overline{\mu}_{PY}$ for the conditional model. Red dashed line: μ_{PY}.

$$\boldsymbol{\Omega} = \left[\begin{array}{c|c|c} \boldsymbol{\Omega}_{PY} & \boldsymbol{\Omega}_{PY,\,CY,s} & \mathbf{0}_{L\times P} \\ \hline & \boldsymbol{\Omega}_{CY,s} & \mathbf{0}_{L\times P} \\ \hline & & \mathbf{I}_{P\times P} \end{array}\right]$$

Table 6.6: Copula correlation matrix from the marginalized model.

LoB	1	2	3	4	5	6	7	8	9
1	1	0.1517	0.1505	0.2501	0.5001	0.2501	0.1501	0.2502	0.2511
2		1	0.152	0.1517	0.1517	0.1517	0.1517	0.1517	0.2532
3			1	0.1505	0.1505	0.1505	0.1505	0.1505	0.2515
4				1	0.2501	0.1501	0.1501	0.1501	0.2511
5					1	0.2501	0.1501	0.2501	0.2511
6						1	0.1501	0.2502	0.2511
7							1	0.1502	0.2511
8								1	0.2511
9									1

Table 6.7: Correlation block for the marginalized model: $\boldsymbol{\Omega}_{PY}$

LoB	1	2	3	4	5	6	7	8	9
1	0.5004	0.5005	0.1502	0.2505	0.2503	0.2504	0.1504	0.2506	0.2506
2	0.5046	0.5046	0.2528	0.1519	0.2528	0.1518	0.1519	0.1519	0.2529
3	0.1506	0.2509	0.5013	0.251	0.1506	0.1506	0.1508	0.1507	0.2511
4	0.2503	0.1502	0.2503	0.5008	0.1502	0.1503	0.1504	0.1504	0.2506
5	0.2503	0.2503	0.1502	0.1503	0.5004	0.2504	0.1504	0.2506	0.2506
6	0.2503	0.1502	0.1502	0.1503	0.2503	0.5006	0.2507	0.2506	0.2506
7	0.1502	0.1503	0.1502	0.1504	0.1502	0.2505	0.501	0.1504	0.2506
8	0.2503	0.1502	0.1502	0.1504	0.2503	0.2504	0.1504	0.5009	0.2506
9	0.2511	0.2511	0.2511	0.2513	0.2511	0.2512	0.2514	0.2513	0.5018

Table 6.8: Correlation block for the marginalized model: $\boldsymbol{\Omega}_{PY,\,CY,s}$

LoB	1	2	3	4	5	6	7	8	9
1	1	0.5006	0.1503	0.2506	0.2504	0.1504	0.1505	0.1505	0.2507
2		1	0.2505	0.1504	0.2504	0.1504	0.1505	0.1505	0.2507
3			1	0.2506	0.1503	0.1504	0.1505	0.1505	0.2507
4				1	0.1504	0.1505	0.1506	0.1506	0.2509
5					1	0.2505	0.1505	0.2507	0.2507
6						1	0.2508	0.2508	0.2508
7							1	0.1507	0.251
8								1	0.2509
9									1

Table 6.9: Correlation block for the marginalized model: $\boldsymbol{\Omega}_{CY,s}$

p_0	1	2	3	4	5	6	7
0.4	0.6	0.84	0.936	0.9744	0.9898	0.9959	
0.5	0.5	0.75	0.875	0.9375	0.9688	0.9844	0.9922
0.7	0.3	0.51	0.657	0.7599	0.8319	0.8824	0.9176

p_0	8	9	10	11	12	13
0.4						
0.5						
0.7	0.9424	0.9596	0.9718	0.9802	0.9862	0.9903

Table 6.10: Intermediate quantiles for different values of p_0.

6.4 Details of the SMC algorithm

6.4.1 Selection of intermediate sets

For both the marginalized and the conditional models we use an adaptive strategy similar to [Cérou et al., 2012] in order to estimate "on the fly" the levels $B_1, \ldots, B_T$. When levels are being chosen "on the fly" one of the main advantages of the proposed SMC algorithm is the ability to estimate, in one run, the company-wide value at risk, the expected shortfall as well as the risk allocations.

Starting from $B_0 = 0$ (or $\overline{B}_0 = 0$ if the conditional model is being used) the idea consists of, at each algorithmic iteration $t-1$, choosing the next level, B_t, such that a percentage $p_0 \in (0,1)$ of the $(t-1)$–particles is above this set. More formally, we set B_t to be the $1-p_0$ empirical quantile of the weighted sample $\{s_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$ or $\{\overline{s}_{t-1}^{(j)}, W_{t-1}^{(j)}\}_{j=1}^N$, where s_{t-1} and $\overline{s}_{t-1}$ denote, respectively, the sum of the components of $\boldsymbol{z}_t$ and $\overline{\boldsymbol{z}}_t$. Therefore, at algorithmic time t the level B_t corresponds to an estimate of the $(1-p_0^t)$-th quantile of the target distribution. In our examples we set $p_0 = 0.4, 0.5$ and 0.7 which induces intermediate quantiles seen in Table 6.10 for the algorithm. Note that, given a value of p_0 the number of levels in the algorithm is deterministic. For example, for $p_0 = 0.5$ there are 7 levels until the estimated quantile is above 99% and one can decide to stop the algorithm at this step or at the previous one, when the quantile level is only slightly lower than 99%.

An alternative approach to choosing the level sets is to use the classic normalizing constant estimator derived from the SMC sampler algorithm (see Section 3.4.2). Using the notation from Section 6.2, we have that the normalizing constant $\mathcal{Z}_t = \mathbb{P}[S > B_t]$ can be estimated as

$$\widehat{\mathcal{Z}}_t = \widehat{\mathcal{Z}}_{t-1} \sum_{j=1}^N W_{t-1}^{(j)} \widetilde{\alpha}_t^{(j)}, \tag{6.12}$$

where W_{t-1} and $\widetilde{\alpha}_t$ are, respectively the normalized and the incremental weights at time $t-1$.

Similarly to our proposed estimate, in this alternative route one would choose B_t such that $p_0 \times 100\%$ of the time $t-1$ particles are above this level. Using the estimator in (6.12) one could stop the algorithm as soon as $\widehat{\mathcal{Z}}_t < \alpha$. The main disadvantage of this approach is that although

$\widehat{\mathcal{Z}}_t$ can be proved to be unbiased and asymptotically normally distributed when the number of particles $N \to +\infty$ (see [Del Moral, 2004, Propositions 7.4.1 and 9.4.1] and [Pitt et al., 2012] for a proof in the special case of state-space models), one can not guarantee that $\widehat{\mathcal{Z}}_t \in [0, 1]$. In our experiments the results based on this classic estimate were deemed unsatisfactory, as we observed estimates of the normalizing constant as large as 15.

6.4.2 Marginalized model

6.4.2.1 The forward kernel

Similarly to Section 4.2.1 we propose a mutation kernel $K_t(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t)$ such that the condition $\sum_{i=1}^{d} z_{i,t} > B_t$ is always satisfied. Due to the independence assumption of the CY large claims (the P Pareto variables) we first independently mutate the Pareto coordinates, following their true (unconditional) marginal and then mutate the other $2L$ variables.

First we split the vector into its log-normal and Pareto components, $\boldsymbol{z}_t = (\boldsymbol{z}'_t, \boldsymbol{z}''_t)$, where $\boldsymbol{z}'_t = (z_{t,1}, \ldots, z_{t,2L})$ and $\boldsymbol{z}''_t = (z_{t,2L+1}, \ldots, z_{t,2L+P})$. Using this notation we use

$$\begin{aligned} K_t(\boldsymbol{z}_{t-1},\, \boldsymbol{z}_t) &= K'_t(\boldsymbol{z}'_{t-1},\, \boldsymbol{z}'_t \,|\, \boldsymbol{z}''_t) \times K''_t(\boldsymbol{z}''_{t-1},\, \boldsymbol{z}''_t) \\ &= \left\{ \frac{1}{2L} \sum_{m=1}^{2L} \left[K_t'^{(-m)}(\boldsymbol{z}'_{t-1}, \boldsymbol{z}'_{t,-m}) K_t'^{(m)}(\boldsymbol{z}'_{t-1}, z'_{t,m} \,|\, \boldsymbol{z}'_{t,-m},\, \boldsymbol{z}''_t) \right] \right\} \\ &\times \prod_{i=2L+1}^{2L+P} Pareto(\boldsymbol{z}''_t;\, \alpha_i,\, \beta_i) \end{aligned}$$

where the kernel $K_t'^{(-m)}(\boldsymbol{z}'_{t-1},\, \cdot\,)$, which mutates all but the m-th dimension of $\boldsymbol{z}'_{t-1}$, consists of independent moves in each dimension, i.e.,

$$K_t^{(-m)}(\boldsymbol{z}'_{t-1}, \boldsymbol{z}'_t) = \prod_{\substack{i=1 \\ i \neq m}}^{2L} K_t'^{(-m,i)}(z'_{t-1,i}, z'_{t,i}).$$

Note that these moves are also independent of the P Pareto mutations.

Let us denote $\{\boldsymbol{z}_{t-1}^{(j)},\, W_{t-1}^{(j)}\}_{j=1}^{N}$ the weighted sample approximating

$$\pi_t(\boldsymbol{z}_{t-1}) = f_{\boldsymbol{Z}}(\boldsymbol{z}_{t-1} \,|\, \boldsymbol{z}_{t-1} \in \mathcal{G}_{\boldsymbol{Z}_{t-1}}),$$

as defined in (6.6). The components of the mutation kernel are then defined as

$$K_t'^{(-m,i)}(z'_{t-1}, z'_{t,i}) = LN(z'_{t,i};\, \widehat{\mu}_i, \widehat{\sigma}_i), \qquad \text{for } i = 1, \ldots, 2L,\ i \neq m \tag{6.13}$$

where $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ are the empirical mean and variance of $\{\boldsymbol{z}_{t-1}^{(j)},\, W_{t-1}^{(j)}\}_{j=1}^{N}$ when $i = 1, \ldots, 2L$

$$\begin{aligned} \widehat{\mu}_{t-1,i} &= \sum_{j=1}^{N} W_{t-1}^{(j)} z_{t-1,i}^{(j)}, \\ \widehat{\sigma}^2_{t-1,i} &= \widehat{\mu}^2_{t-1,i} - \sum_{j=1}^{N} W_{t-1} \left(z_{t-1,i}^{(j)} \right)^2. \end{aligned}$$

For the mutation of the remaining dimension, m, to ensure all the samples satisfy the condition $\sum_{i=1}^{d} z_{i,t} > B_t$ we proceed as follows. First we define

$$B_t^z(m) = \max\left\{0,\, B_t - \sum_{\substack{i=1\\ i\neq m}}^{d} z_{t,i}\right\}$$

and then sample the last component $z_{m,t} \in [B_t^z(m), +\infty)$ according to

$$K_t^{(m)}(\boldsymbol{z}_{t-1}, z_{t,m} \,|\, \boldsymbol{z}_{t,-m}) = TN(z_{t,m};\, \widehat{\mu}_m, \widehat{\sigma}_m, B_t^z(m), +\infty), \quad \text{for } m = 1, \ldots, 2L, \tag{6.14}$$

where $TN(\,\cdot\,; \mu, \sigma, a, b)$ denotes the density of a Normal with mean μ and variance σ^2 truncated at $[a, b]$.

6.4.2.2 The backward kernel

For the backward kernel we follow the discussion in Section 3.4.3.1 and use the (approximation to the) optimum kernel, given by equation (3.12)

$$L_t(\boldsymbol{z}_{t+1}, \boldsymbol{z}_t) = \frac{\gamma_t(\boldsymbol{z}_t) K_{t+1}(\boldsymbol{z}_t, \boldsymbol{z}_{t+1})}{\frac{1}{N}\sum_{j=1}^{N} w_t^{(j)} K_{t+1}(\boldsymbol{z}_t^{(j)}, \boldsymbol{z}_{t+1})},$$

where $w_t^{(j)}$ denotes the *unnormalized* weights at time t and the weighted sample $\{z_t^{(j)},\, w_t^{(j)}\}_{j=1}^{N}$ targets the unnormalized density $\gamma_t(\boldsymbol{z}_t)$. Proceeding this way the unnormalized weights for the SMC sampler algorithm (see Algorithm 8) satisfy the following recursion

$$w_t^{(j)} = w_{t-1}^{(j)} \frac{\gamma_t(\boldsymbol{z}_t)}{\frac{1}{N}\sum_{k=1}^{N} w_t^{(k)} K_t(\boldsymbol{z}_{t-1},\, \boldsymbol{z}_t)}.$$

6.4.2.3 The MCMC move kernel

To improve particle diversity after a resampling step (which is performed whenever the effective sample size drops bellow $N/2$) the following MCMC move kernel is applied to the particles.

As in Chapter 4 we propose a Gibbs-type update combined with a slice sampler. For notational simplicity we suppress the dependence in t in the vector $\boldsymbol{z}_t$ and denote $v^*(m) = (z_1^*, \ldots, z_m^*, z_{m+1}, \ldots, z_d)$ the vector where the first m components have already been updated in the Gibbs scan. The full conditional for the m-th component of $\boldsymbol{z}_t$ is given by

$$\pi_t(z_m^* \,|\, z_1^*, \ldots, z_{m-1}^*, z_{m+1}, \ldots, z_d) \propto \pi_t(v^*(m)) \propto f_{\boldsymbol{Z}}(v^*(m)) \mathbb{1}_{\mathcal{G}_{\boldsymbol{Z}_t}}(v^*(m)),$$

which can be sampled from using an unidimensional slice sampler.

6.4.3 Conditional model

Following the discussion in Section 6.2.2.2 we use equation (6.10) as an approximation to the unknown density $f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$. For our simulations $M = 5$ samples of the unknown parameter $\boldsymbol{\theta}$ are used, where

$$\boldsymbol{\theta} = (\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(L)})$$

and each vector $\boldsymbol{\sigma}^{(\ell)} = (\sigma_1^{(\ell)}, \ldots, \sigma_J^{(\ell)})$ contains all the unknown variance parameters for the ℓ-th LoB. Therefore, $\boldsymbol{\vartheta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)})$ and it should be noticed the superscript has a different interpretation from those in $\sigma_j^{(\ell)}$.

As the parameter estimation step described in Section 6.3.4 is independent of the allocation process we assume N_{MCMC} samples for each unknown parameter vector $\boldsymbol{\sigma}$ have already been generated. Therefore, to sample $\overline{\boldsymbol{z}} \sim f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}})$ we first sample an index $n \sim U(\{1, \ldots, N_{MCMC}\})$ and then $\overline{\boldsymbol{z}} \sim f_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}} \,|\, \boldsymbol{\theta}^{(n)})$.

6.4.3.1 The forward kernel

The forward kernel used for the conditional model follows the same structure as the one used in the marginalized model and described in Section 6.4.2.1: first we sample the P independent Pareto variables (with the same distribution as in the marginalized case) and then the remaining $2L$ variables. More precisely,

$$\overline{K}_t'^{(-m,i)}(\overline{\boldsymbol{z}}'_{t-1},\, \overline{z}'_{t,i} \,|\, \boldsymbol{\vartheta}_t) = \overline{K}_t'^{(-m,i)}(\overline{\boldsymbol{z}}'_{t-1},\, \overline{z}'_{t,i}) = K_t'^{(-m,i)}(\overline{\boldsymbol{z}}'_{t-1}, \overline{z}'_{t,i})$$

where the last term is defined in equation (6.13) and $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ are now the empirical mean and variance of $\{\overline{\boldsymbol{z}}_{t-1}^{(j)},\, W_{t-1}^{(j)}\}_{j=1}^N$. Likewise,

$$\overline{K}_t'^{(m)}(\overline{\boldsymbol{z}}'_{t-1}, \overline{z}'_{t,m} \,|\, \overline{\boldsymbol{z}}'_{t,-m},\, \boldsymbol{\vartheta}_t) = \overline{K}_t'^{(m)}(\overline{\boldsymbol{z}}'_{t-1}, \overline{z}'_{t,m} \,|\, \overline{\boldsymbol{z}}'_{t,-m}) = K_t'^{(m)}(\overline{\boldsymbol{z}}'_{t-1}, \overline{z}'_{t,m} \,|\, \overline{\boldsymbol{z}}'_{t,-m})$$

with the last term defined in equation (6.14). As samples from $f_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta})$ have already been generated through MCMC then the mutation kernel in the extended space, $K_t^{\mathbf{y}}(\mathbf{y}_{t-1}, \mathbf{y}_t)$, is completely characterized.

6.4.3.2 The backward kernel

As in Section 6.4.2.2 we use the optimum backward kernel in the extended space $\mathcal{Y} = \mathbb{R}^d \times \Theta^M$, which for the conditional model leads to the following incremental weights (see equation (3.13))

$$\begin{aligned}
\alpha_t &= \frac{\gamma_t^{\mathbf{y}}(\mathbf{y}_t)}{\frac{1}{N}\sum_{j=1}^N w_{t-1}^{(j)} K_t^{\mathbf{y}}(\mathbf{y}_{t-1}, \mathbf{y}_t)} \\
&= \frac{\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t;\, \boldsymbol{\vartheta}_t) f_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t)}{\frac{1}{N}\sum_{j=1}^N w_{t-1}^{(j)} K_t(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) f_{\vartheta}(\vartheta_t)} \\
&= \frac{\widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t;\, \boldsymbol{\vartheta}_t) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t)}{\frac{1}{N}\sum_{j=1}^N w_{t-1}^{(j)} K_t(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t)}.
\end{aligned}$$

6.4.3.3 The MCMC move kernel

The MCMC move kernel used for the conditional model needs to keep the target distribution in the extended space, $\pi_t^{\mathbf{y}}(\mathbf{y}_t)$, invariant. The strategy adopted is to first sample $\boldsymbol{\vartheta}^* \sim f_{\vartheta}(\vartheta)$ and then $\boldsymbol{z}_t \,|\, \boldsymbol{\vartheta}^* \sim \widehat{f}_{\overline{\boldsymbol{Z}}}(\overline{\boldsymbol{z}}_t;\, \boldsymbol{\vartheta}_t^*) \mathbb{1}_{\mathcal{G}_{\overline{\boldsymbol{Z}}_t}}(\overline{\boldsymbol{z}}_t)$.

For the second step above we use exactly the same Gibbs-sampler update as in Section 6.4.2.3, with $f_{\boldsymbol{Z}}(\,\cdot\,)$ replaced by $\widehat{f}_{\overline{\boldsymbol{Z}}}(\,\cdot\,;\, \boldsymbol{\vartheta}_t)$.

6.5 Results

In this section we present the results of the SMC procedure when used to calculate the marginal expected shortfall allocations from (6.2) and (6.3) and, consequently, the solvency capital requirement.

Before proceeding to the results calculated via the SMC algorithm, in order to understand the simulated data presented in Figure 6.1, in Table 6.5 we present some results based on a "brute force" Monte Carlo (rejection-sampling) simulation, which is taken as the base line for comparisons with the SMC algorithm. Table 6.5 is divided in three blocks of rows, with PY claims, CY small (CY,s) claims and CY large (CY,l) claims.

First of all, it should be noticed that the reserves presented on the first block of Table 6.5 are the ones implied by the data, which we then assume to be the true ones (ignoring the original data from Table 6.1). The parameters σ and μ for PY claims are related to the marginalized model (for the parameters of the conditional model see Figures 6.5 and 6.6). It is also important to note that only the PY parameters are different between the conditional and marginalized models.

For each LoB the standalone expected shortfall (ES) is calculated analytically and its value is, then, combined with the LoB's expectation to calculate the solvency capital requirement (SCR). These values are added up, both within risk type (i.e., PY, CY,s and CY,l) and globally, in order to calculate the overall standalone capital. For the marginalized and conditional models the columns "ES" and "SCR" denote, respectively, the expected shortfall and capital allocations to each LoB. These values are compared to its standalone counterparts to generate the diversification benefit, which is around 45% for PY and CY,s claims (regardless of the model used) and ranges between 30% and 70% within the PY and CY,s groups. Due to the independence assumption the largest diversification benefit comes from the CY,l claims, where the capital is reduced by around 95%.

The data presented in Table 6.5 is calculated as follows. For the marginalized model (and conditional model in brackets), 5×10^9 (2.5×10^7) independent samples of the model are generated in order to calculate the overall $\text{VaR}_{99\%}$. Conditional on this value, for each LoB we then generate 5×10^7 (5×10^5) samples above the VaR and use the average of these samples as the true ES allocation (presented in Table 6.5). In order to asses the variance of the estimators, we divide these samples into $N_{rep} = 500$ groups of $N_{MC} = 10^5$ ($N_{MC} = 10^3$ for the conditional model) simulations. More formally, we approximate the ES allocations $\mathcal{A}_i$, defined in (6.2), by

$$\widehat{\mathcal{A}}_{i,MC} = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \widehat{\mathcal{A}}_{i,MC}^{(k)}, \tag{6.15}$$

where $\widehat{\mathcal{A}}_{i,MC}^{(k)}$ stands for the estimate (using N_{MC} particles) from the k-th run (out of N_{rep}), and is defined as

$$\widehat{\mathcal{A}}_{i,MC} = \frac{1}{N_{MC}} \sum_{j=1}^{N_{MC}} Z_i^{(j)}.$$

Similarly to the analysis performed in [Peters et al., 2017] the impact of the prior density can be assessed by comparing the sum of the SCR allocations with the SCR from the "empirical Bayes model", i.e., the model where the prior for $\boldsymbol{\sigma}$ is set as a Dirac mass on $\widehat{\sigma}_j(t)$, see (6.11). In this case we have that the total capital is equal to SCR = 505.48 and the fully Bayesian

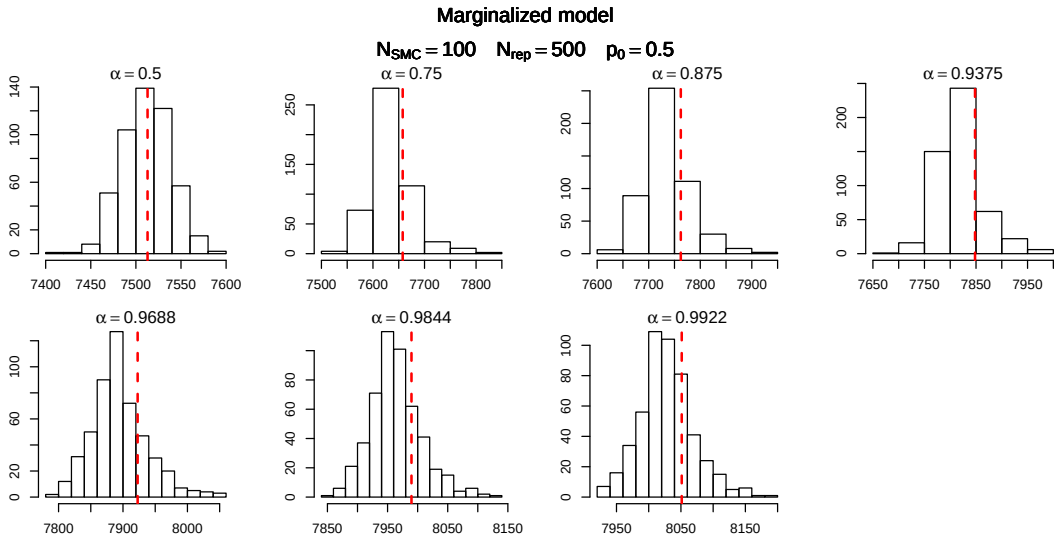

Figure 6.7: Histograms levels used in the SMC sampler algorithm with $p_0 = 0.5$ in the marginalized model. The red dashed bar represents the true value of the α quantile.

model with prior defined with $k = 5$ (see Section 6.3.4) requires 15% more capital (both in the marginalized and conditional cases).

To check the accuracy of the SMC procedure we first analyse the estimate of the level sets (intermediate VaRs). For $p_0 = 0.5$, Figures 6.7 and 6.8 show, respectively, the histogram of the levels $B_1, \ldots, B_7$ (as per Table 6.10) for the marginalized and conditional models. The red dashed bars represent the true value of the quantiles (based on the "brute force" MC simulations), which is very close to the mode of the empirical distribution of the SMC estimates. It should be noticed, though, that the SMC estimates seem to be negatively biased and the bias appears to become more pronounced for extreme quantiles. Apart from this negligible bias we assume the levels are being sensibly estimated and proceed, as in Chapter 4, to calculate the relative bias and the variance reduction of the SMC method when compared to a MC procedure.

For each of the LoBs the plots on the Figures 6.9 and 6.10 show the relative bias, defined as

$$\text{Relative Bias} = \frac{\widehat{\mathcal{A}}_{i,SMC} - \widehat{\mathcal{A}}_{i,MC}}{\widehat{\mathcal{A}}_{i,MC}},$$

where $\widehat{\mathcal{A}}_{i,SMC}$ is computed analogously to the MC estimate but, instead, using the SMC method, with $N_{SMC} = 100$. The behaviour of the two models is very similar, and we observe that the bias in the PY and CY,s allocations are negligible (less than 5%) while for some of the large CY risks a higher bias (of more than 10%) may be observed. Apart from the difficulty of performing the estimation based on Pareto distributions we stress the fact that although these errors may look large, as we can see from Table 6.5, their impact in the overall capital are almost imperceptible, due to the small capital charge due to these risks.

Another way to compare the SMC calculations is through the actual capital charges, as seen in Figures 6.11 and 6.12. In these figures we compare the 99% SCR calculated via the MC

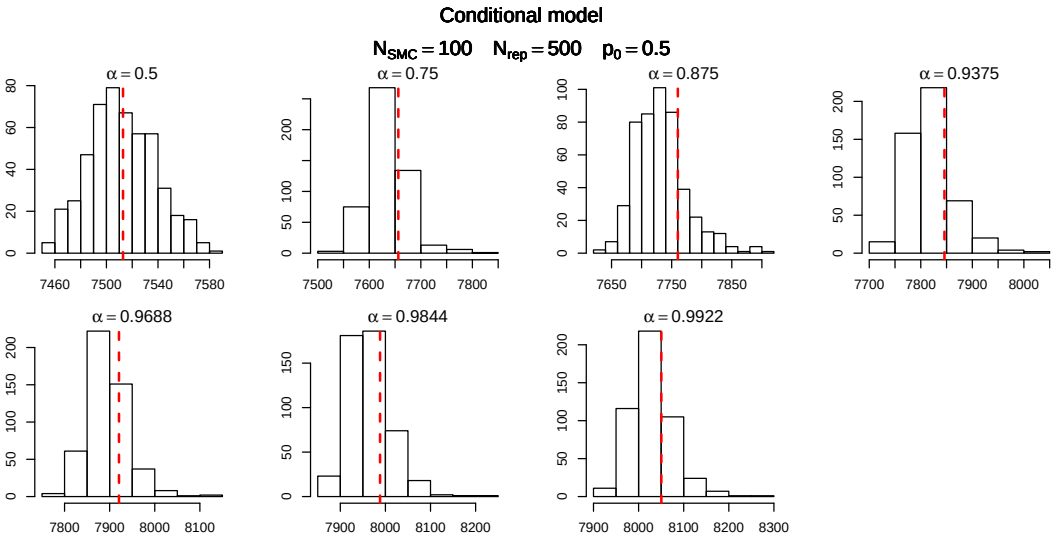

Figure 6.8: Histograms levels used in the SMC sampler algorithm in the conditional model. The red dashed bar represents the true value of the α quantile.

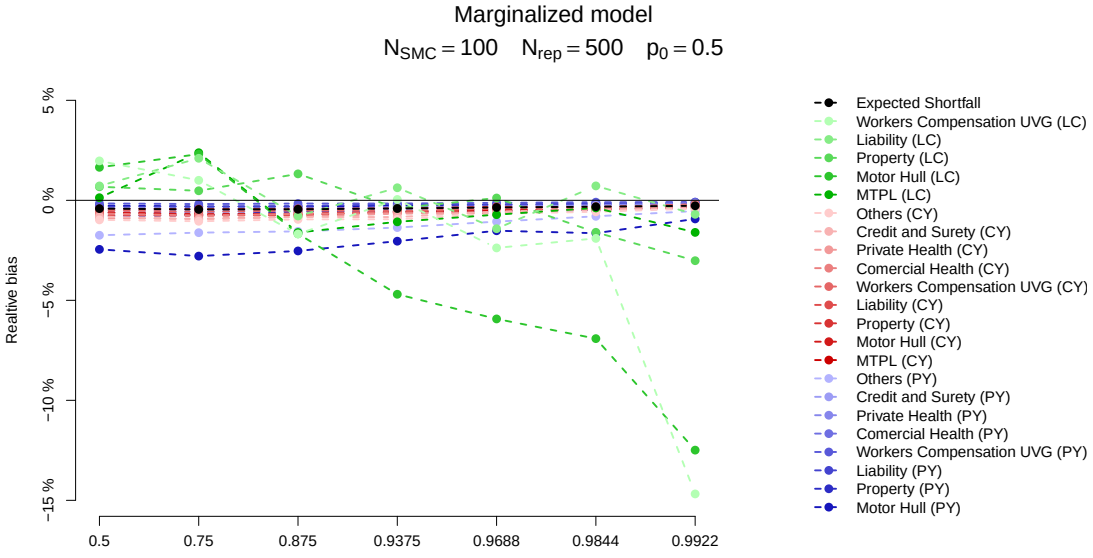

Figure 6.9: Bias for the marginalized model.

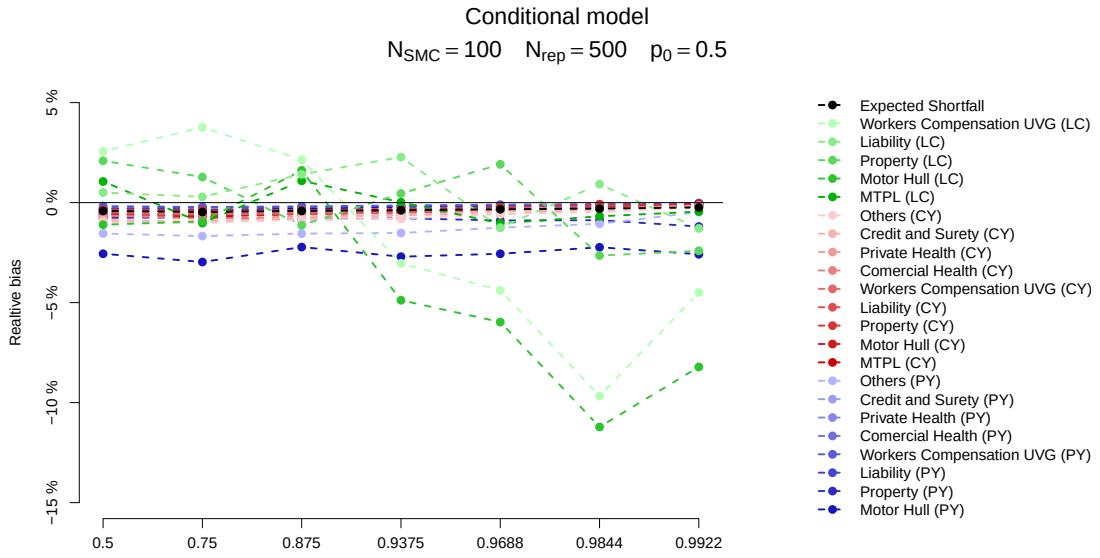

Figure 6.10: Bias for the conditional model.

scheme discussed above with the SMC results for the quantile level right before 99% (which, for $p_0 = 0.5$ is 98.44%) and the one right after it (99.22%). From these figures we see the SMC calculation based on the 99.22% quantile is very precise, for both the marginalized and conditional models. Visually, the only perceivable difference comes from the CY,l claims, which accounts (in total) for less than 2% of the overall capital.

To calculate the improvement generated by the SMC algorithm compared to the MC procedure we need to analyse the variance of the estimates generated by both methods, under similar computational budgets.

We start by noticing that the expected number of samples in the Monte Carlo scheme in order to have N_{MC} samples satisfying the α condition is equal to $M_{MC} = N_{MC}/(1-\alpha)$, which can be prohibitive if α is very close to 1. Then, similarly to (6.15) we define the empirical variance of the MC and the SMC algorithms which are, then, compared as follows

$$\text{Variance Reduction} = M_{MC} \times \widehat{\text{Var}}(\widehat{\mathcal{A}}_{i,MC}) \Big/ T \times N_{SMC} \times \widehat{\text{Var}}(\widehat{\mathcal{A}}_{i,SMC}). \tag{6.16}$$

The variance reduction statistic defined in (6.16) takes into account how many samples one needs to use in order to generate N_{MC} samples via rejection sampling or N_{SMC} using the SMC algorithm. The later also takes into account the fact that T levels are being used and in each one N_{SMC} samples need to be generated. For the conditional model we further multiply the denominator by the number of samples used to estimate the unknown density, which in our examples is set to $M = 5$.

The results follow on Figures 6.13 and 6.14. As in Chapter 4 we observe that the variance of the SMC estimates become smaller (compared to the MC results) for larger quantiles. In particular, for the quantiles of interest the variance of the marginal ES allocation estimates are

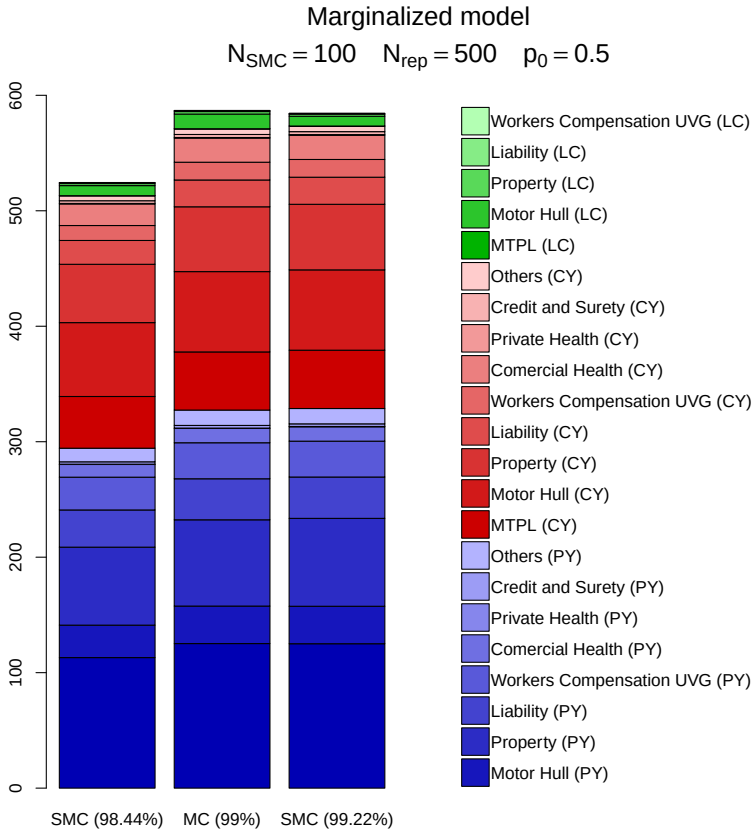

Figure 6.11: Comparison between the "true" allocations (calculated via a large Monte Carlo procedure) and the SMC sampler solution for the marginalized model.

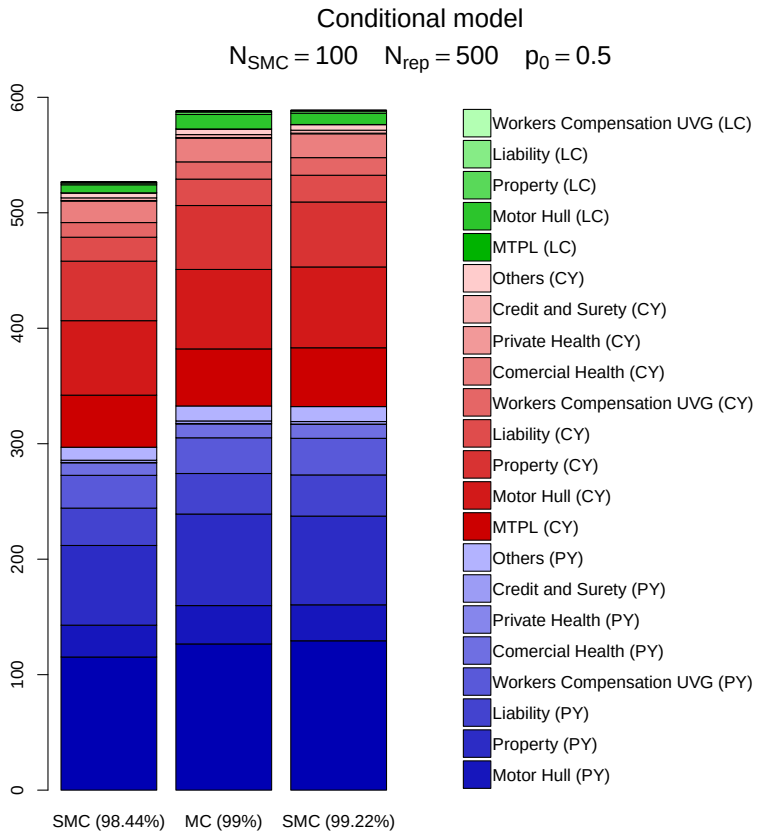

Figure 6.12: Comparison between the "true" allocations (calculated via a large Monte Carlo procedure) and the SMC sampler solution for the conditional model.

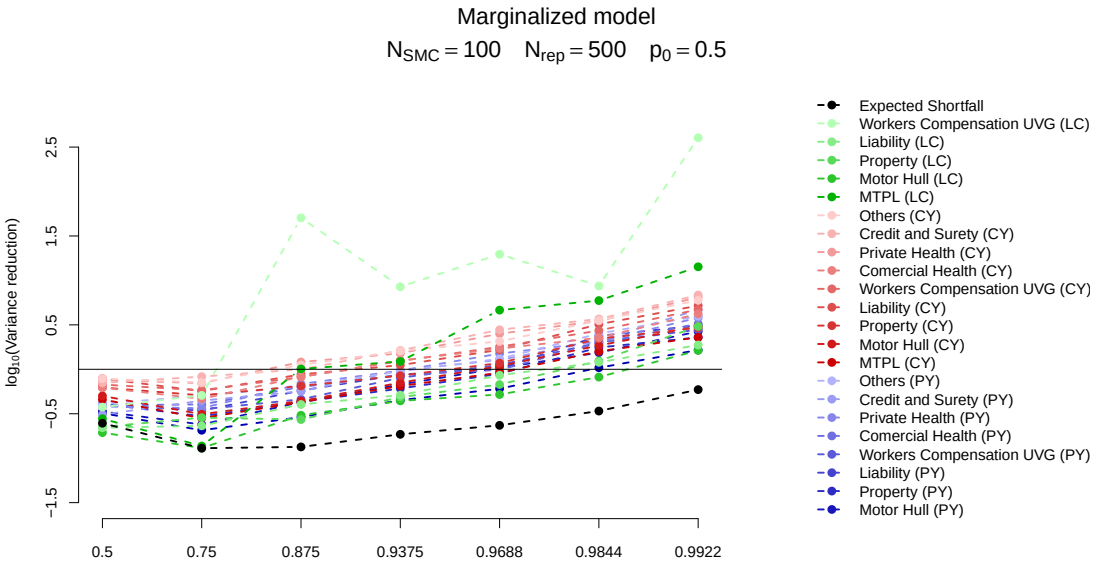

Figure 6.13: Variance reduction for the marginalized model.

around 3 times smaller than its MC counterparts, while the overall ES estimate is slightly less variable for the MC scheme.

For the marginalized model we also present two plots, Figures 6.15 and 6.16, related, respectively, to the sensitivity to the parameter p_0 and to the number of samples, N_{SMC}. In Figure 6.15, for the same number of samples, $N_{SMC} = 100$ we analyse the bias relative to the 99% ES allocations of the first quantile larger than 99% (top plot) and the previous one (bottom plot) for $p_0 \in \{0.4, 0.5, 0.7\}$. The quantiles used in these different setups are presented in Table 6.10. Although the results may look slightly different the main message is the same: the "higher" quantile is effectively unbiased for PY and CY,s risks but presents a negative bias of around 10% for some of the CY,l risks.

Regarding the sensitivity to the number of particles in the SMC algorithm, as expected, the absolute bias decreases when the number of samples increases, as seen in Figure 6.16. Although the SMC algorithm is generically guaranteed to be unbiased when $N_{SMC} \to +\infty$ the trade-off between bias and the variance reduction in the allocation problem may lead us to accept a small bias in order to have a smaller variance.

6.6 Conclusions

In this chapter we provide a complete and self-contained view of the capital allocation process for non-life insurance companies. As prescribed by the Swiss Solvency Test we break down the company's overall Solvency Capital Requirement (SCR) into the one-year reserve risk, due to claims from previous years (PY) and the one-year premium risk due to claims' payments in the current year (CY). The later is further split into the risk of normal/small claims (CY,s) and large claims (CY,l). For the premium risk in each line of business we assume a log-normal distribution

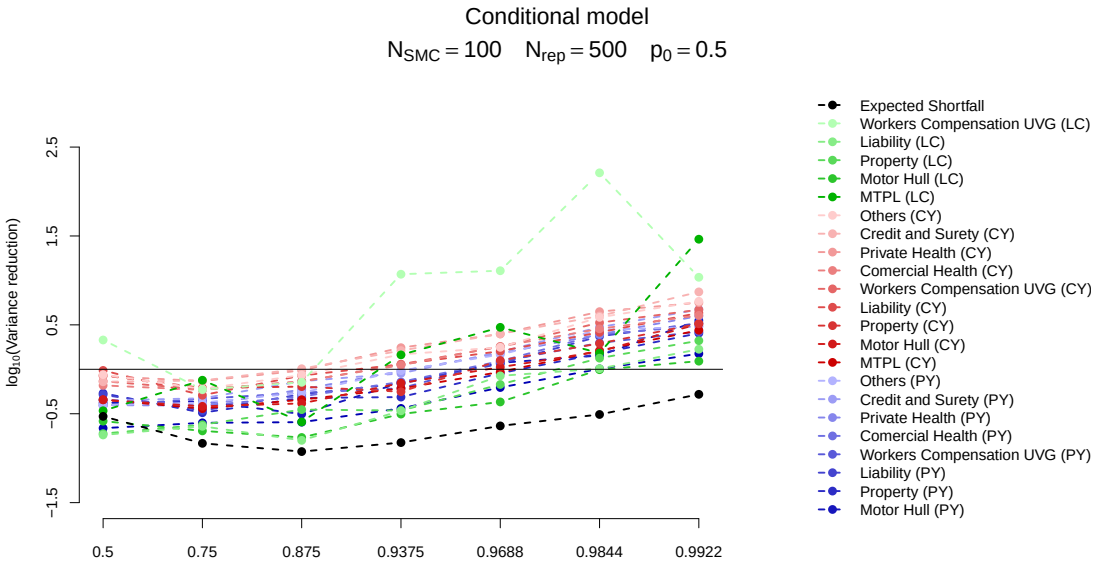

Figure 6.14: Variance reduction for the conditional model.

for CY,s risks with mean and variance as per the SST, which also describe a distribution for CY,l risks, in this case Pareto. For the reserve risk, as in [Peters et al., 2017], we postulate a Bayesian gamma-gamma model which, for allocation purposes, is approximated by log-normal distributions leading to what we name the marginalized and the conditional models.

The allocation process is performed using state-of-the-art (pseudo-marginal) Sequential Monte Carlo (SMC) algorithms, which are presented in a self-contained and accessible format. Although the algorithms described form an extremely flexible class, we provide an off-the-shelf version, where minimal or no tuning is needed. The algorithms are also shown to be computationally efficient in a series of numerical experiments.

One of the advantages of our proposed methodology is that it is able to compute in one single *loop* (1) the value at risk (VaR) and (2) the Expected Shortfall (ES), both at the company level and (3) the capital allocations for the risk drivers. This procedure should be compared with routinely applied methodologies, where one simulation is performed to compute the VaR, which is used in a different simulation to compute the ES and only then a final simulation uses these two estimates to calculate the allocations, in a process that accumulate different errors.

Moreover, even ignoring the computational cost of calculating a precise estimate for the required VaR in a "brute force" Monte Carlo scheme, the proposed SMC algorithm is numerically shown to provide estimates that are less volatile than comparable "brute force" implementations.

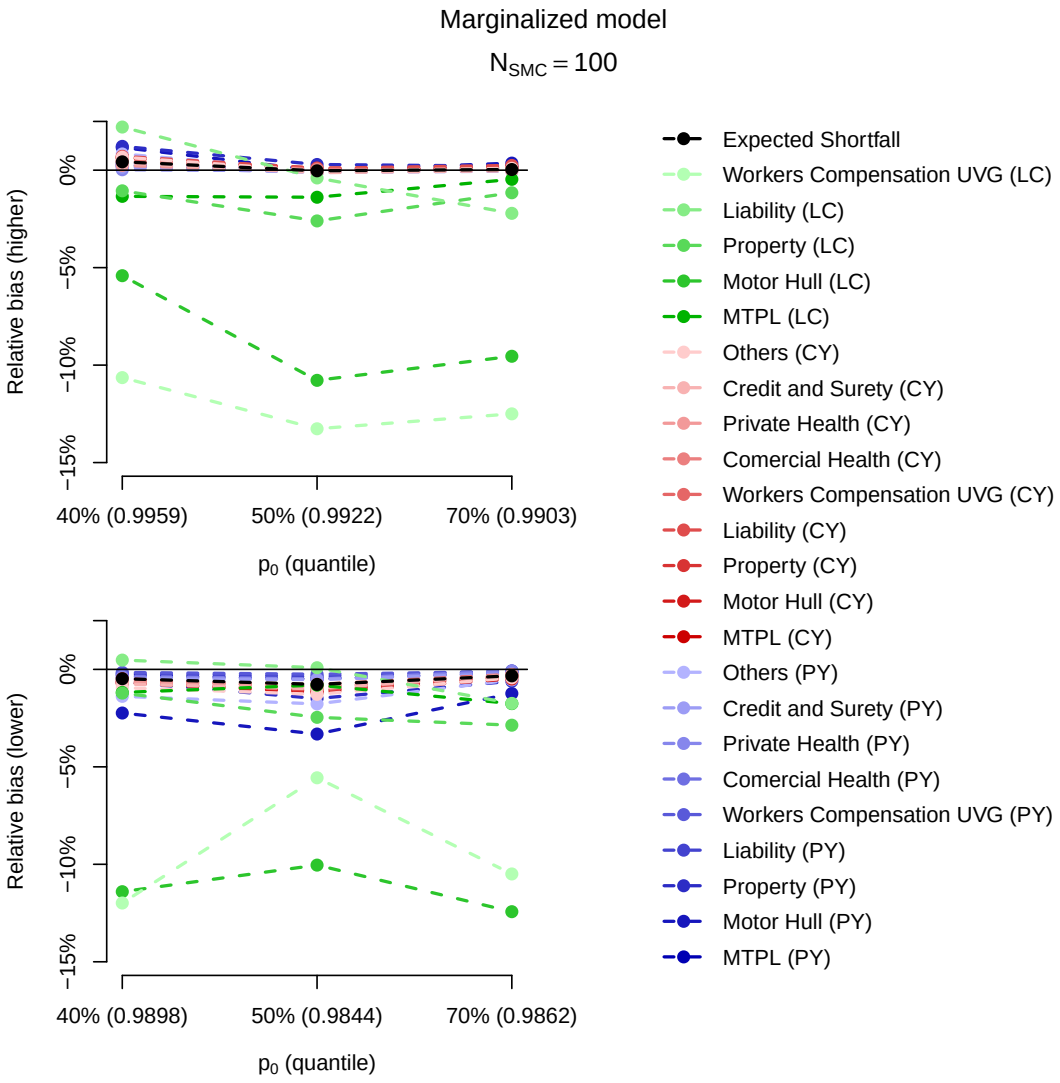

Figure 6.15: Relative bias in the marginalized model as a function of the parameter p_0.

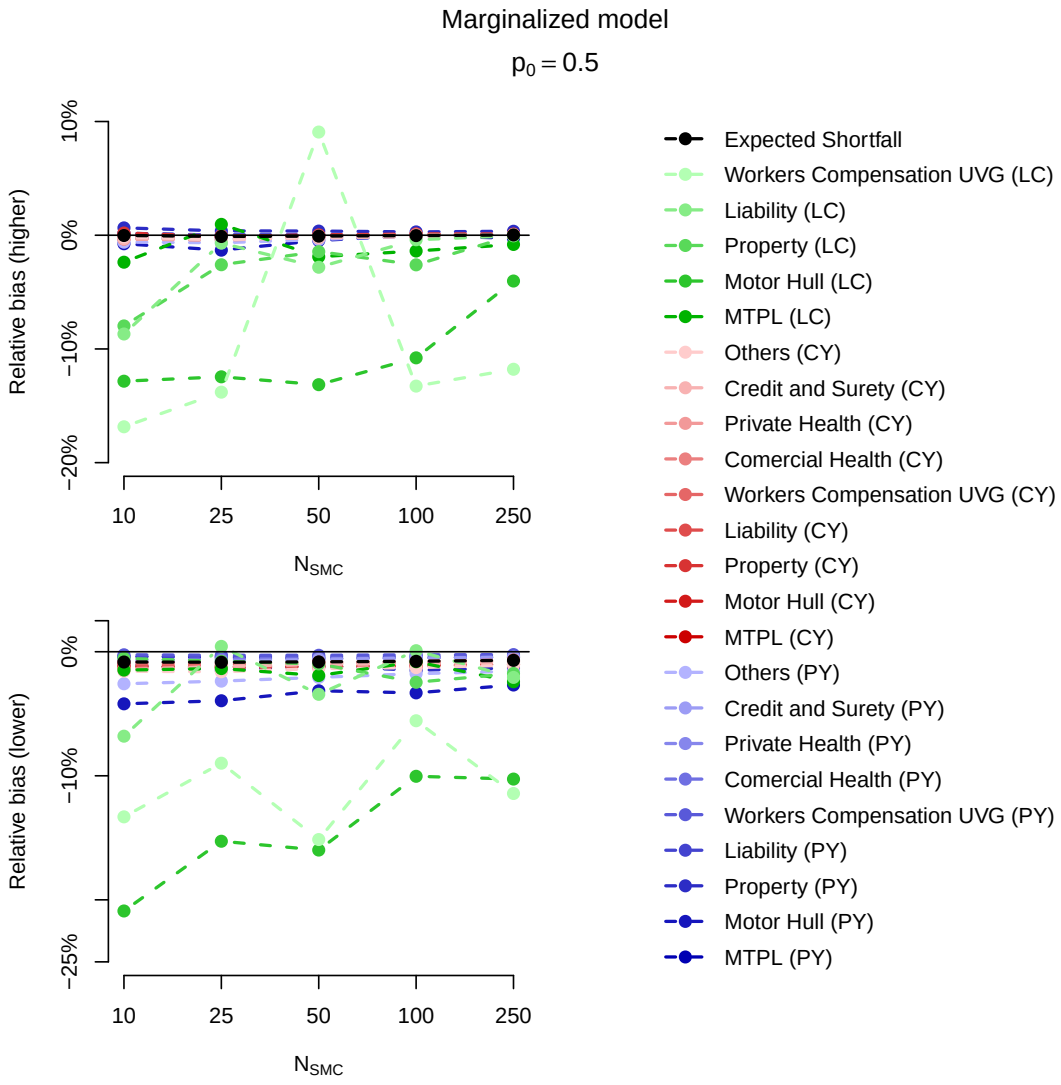

Figure 6.16: Relative bias in the marginalized model as a function of the sample size in the SMC sampler, N_{SMC}.

Chapter 7

Multiple optimal stopping times

In this chapter we return the focus to operational risk modelling and propose a novel class of insurance products that can help financial companies to offload some of its operational losses. The class of products introduced in this chapter lasts for T years but the policyholder only has the right to make claims on $k < T$ years, which can make it affordable to a wider class of buyers. For two different types of annual insurance coverage we derive closed form strategies for the optimal usage of the product (i.e., when to use the k rights). When the combination of loss distribution and insurance mitigation does not lead to closed form usage strategies we also derive analytic expressions based on series expansions. The results of this chapter were published in [Targino et al., 2016].

7.1 Introduction

Since the New Basel Capital Accord in 2004, Operational Risk (OpRisk) quantification has become increasingly important for financial institutions. However, the same degree of attention has not yet been devoted to insurance mitigation of OpRisk losses nor, consequently, to detailed analysis of potential risk and capital reduction that different risk transfer strategies in OpRisk may allow.

Historically the transference of credit and market risks through credit derivatives and interest rate swaps, for example, has been an active subject of extensive studies both from practitioners and academics while only a few references about OpRisk transfer of risk and possible approach to such risk transfers can be found in the literature (see [Brandts, 2004], [Bazzarello et al., 2006] and [Peters et al., 2011]). In the banking industry, Credit Suisse has very recently issued around CHF 220 millions worth of bonds said to cover OpRisk losses between CHF 3.5 billions and CHF 4.2 billions (see [Das and Scism, 2016] and [Foerster and Beardsworth, 2016]), in one of the first operations of this kind.

This slow uptake of insurance policies in OpRisk for capital mitigation can be partially attributed to four general factors: (a) there still remains a rather limited understanding of the impact on capital reduction of currently available OpRisk insurance products, especially in the complex multi-risk, multi-period scenarios; (b) the relative conservative Basel II regulatory

cap of 20% in a given year (for Advanced Measurement Approach models); (c) the limited understanding at present of the products and types of risk transfer mechanisms available for OpRisk processes; and (d) the limited competition for insurance products available primarily for OpRisk, where yearly premiums and minimum Tier I capital requirements required to even enter into the market for such products precludes the majority of banks and financial institutions in many jurisdictions.

Some of the reasons for these four factors arises when one realizes that OpRisk is particularly challenging to undertake general risk transfer strategies for, since its risk processes range from loss processes which are insurable in a traditional sense to infrequent high consequence loss processes which may be only partially insurable and may result from extreme losses typically covered by catastrophe bonds and other types of risk transfer mechanisms. For these reasons, the development of risk transfer products for OpRisk settings by insurers is a relatively new and growing field in both academic research and industry, where new products are developed as greater understanding of catastrophe and high consequence low frequency loss processes are better understood.

It is noted in [Chernobai et al., 2008] that the existence of such specialised products is limited in scope and market since the resulting premium one may be required to pay for such an insurance product can typically run into very significant costs, removing the actual gain from obtaining the insurance contract in terms of capital mitigation in the first place. Hence, although the impact of insurance in OpRisk management is yet to be fully understood it is clear that it is a critical tool for the management of exposures and should be studied more carefully.

In this chapter we discuss aspects of an insurance product that provides its owner several opportunities to decide which annual OpRisk loss(es) to insure. This product can be thought of as a way to decrease the cost paid by its owner to the insurance company in a similar way to what occurs with swing options in energy markets (see for example, [Jaillet et al., 2004] and [Carmona and Touzi, 2008]): *instead of buying T yearly insurance policies over a period of T years, the buyer can negotiate with the insurance company a contract that covers only k of the T years (to be chosen by the owner).* This type of structured product will result in a reduction in the cost of insurance or partial insurance for OpRisk losses and this aspect is highlighted in [Allen et al., 2009, page 188], where they note that "even without considering the cost of major catastrophes, insurance coverage is very expensive". In addition, we argue it may be interesting to explore such structures if the flexibility they provide results in an increased uptake of such products for OpRisk coverage, further reducing insurance premiums and resulting perhaps in greater competition in the market for these products.

The general insurance product presented here can accommodate any form of insurance policy, but we focus on two generic "building block" policies (see Definitions 7.2.1 to 7.2.2) which can be combined to create more complex types of protection. For these two basic policies we present a "moderate-tailed" model for annual risks that leads to closed form usage strategies

of the insurance product, answering the question: when is it optimal to ask the insurance company to cover the annual losses?

For the rest of the chapter we assume that throughout a year a financial institution incurs a random number of loss events, say N, with severities (loss amounts) $X_1, \ldots, X_N$. Additionally, we suppose the company holds an insurance product that lasts for T years and grants the company the right to mitigate k of its T annual losses through utilisation of its insurance claims. To clarify consider a given year $t \leq T$ where the company incur $N(t)$ losses adding up to $Z(t) = \sum_{n=1}^{N(t)} X_n(t)$, assuming it has not yet utilised all its k insurance mitigations it then has the choice to make an insurance claim or not. If it utilises the insurance claim in this year the resulting annual loss is denoted by $\widetilde{Z}(t)$. Such a loss process model structure is standard in OpRisk and insurance and is typically referred to as the Loss Distributional Approach (LDA) which we illustrate an example instance of in Figure 7.1.

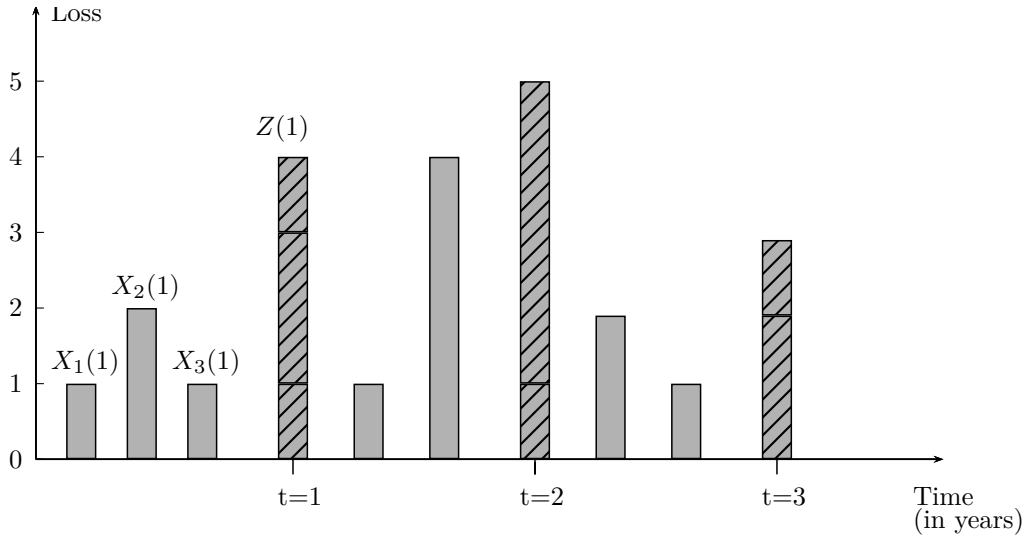

Figure 7.1: Schematic representation of a LDA model. The aggregated loss in each year is represented hatched.

In this context the company's aim is to choose k distinct years out of the T in order to minimize its expected operational loss over the time interval $[0, T]$, where it is worth noting that if $Z > \widetilde{Z}$ i.e., if the insurance is actually mitigating the company's losses, all its k rights should be exercised. Then the question that must be addressed is what is the optimal decision rule, or, in other words, how to define the multiple optimal stopping times for making the k sets of insurance claims.

The rest of the chapter is organized as follows. In Section 7.2 we present the insurance policies we use as mitigation for the insurance product described above. Section 7.3 presents an overview of useful theoretical results in the field of multiple stopping rules for independent observations in discrete time, in particular we state Theorem 7.3.5 (proved in [Nikolaev and Sofronov, 2007]) which is the main result in this section. A summary of prop-

erties related to the LDA model used in this chapter is presented in Section 7.4 and used in Section 7.5 to present the main contribution of this work, namely closed form solutions for the optimal multiple stopping rules for the insurance products considered. In Section 7.6 we check the theoretical optimality of the rules derived in Section 7.5, comparing them with predefined rules.

Since these closed form results rely upon the stochastic loss model considered, we also provide a general framework applicable for any loss process. In Section 7.7 we discuss a method based on series expansions of unknown densities to calculate the optimal rules when the combination of insurance policy and severity density does not lead to analytical results. The conclusions and some final considerations are shown in Section 7.8.

7.2 Insurance policies

As previously mentioned, the insurance policies presented here must be thought as building blocks for more elaborated ones, leading to mitigation of more complex sources of risk. It also worth noticing that the policies presented are just a mathematical model of the actual policies that would be sold in practice and although some characteristics, such as deductibles, can be incorporated in the model they are not presented at this stage.

In the sequel we present these basic insurance policies a company can use within the insurance product. For the sake of notational simplicity, if a process $\left\{Z(t)\right\}_{t=1}^{T}$ is a sequence of i.i.d. random variables, we will drop the time index and denote a generic r.v. from this process as Z. For the rest of the chapter $\mathbb{1}_A$ will denote the indicator function on the event A, ie, $\mathbb{1}_A = 1$ if A is valid and zero otherwise.

Definition 7.2.1 (Individual Loss Policy (ILP))**.** *This policy applies a constant haircut to the loss process in year t in which individual losses experience a Top Cover Limit (TCL) as specified by*

$$\widetilde{Z} = \sum_{n=1}^{N} \max\left(X_n - TCL,\ 0\right).$$

Definition 7.2.2 (Accumulated Loss Policy (ALP))**.** *The ALP provides a specified maximum compensation on losses experienced over a year. If this maximum compensation is denoted by ALP then the annual insured process is defined as*

$$\widetilde{Z} = \left(\sum_{n=1}^{N} X_n - ALP\right) \mathbb{1}_{\left\{\sum_{n=1}^{N} X_n > ALP\right\}}.$$

To characterize the annual application of such policies we provide a schematic representation of each of these policies in Figures 7.2 to 7.3, assuming the same losses as in Figure 7.1. The (part of the) loss mitigated by the insurance policy is represented by a white bar and the remaining loss due to the owner of the insurance product is coloured grey. As in Figure 7.1, annual losses are represented by hatched bars.

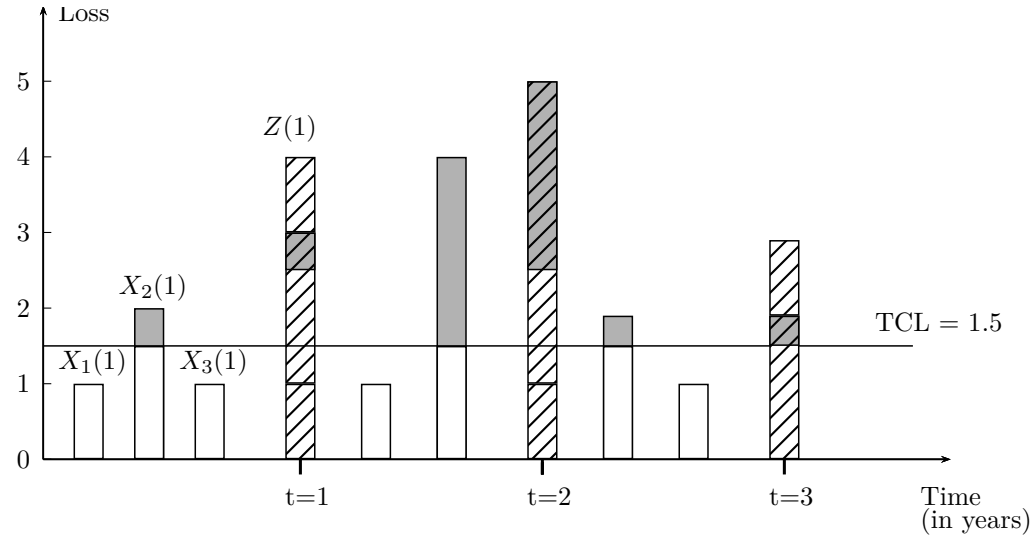

Figure 7.2: Individual Loss Policy (ILP) with TCL level of 1.5.

7.3 Multiple optimal decision rules

Assume an agent sequentially observe a process $\{W(t)\}_{t=1}^{T}$, for a fixed $T < +\infty$ and wants to choose $k < T$ of these observations in order to maximize (or minimize, see Remark 7.3.7) the expected sum of these chosen observations. For $k = 1$, this problem is known in the literature as the house selling problem (see [Sofronov, 2013] for an updated literature review) since one of its interpretations is as follows. If the agent is willing to sell a house and assume that at most T bids will be observed he wants to choose the optimal time τ such that the house will be sold for the highest possible value. The extension of this problem for $k > 1$ is know as the multiple house selling problem, where the agent wants to sell k identical houses. It is worth noting that in our insurance problem the agent is interested in choosing k periods to exercise the insurance policy in order to minimize its loss, in a sense that will be make precise shortly in this chapter.

Formally, the mathematical framework of this problem consists of a filtered probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\big)$, where $\mathcal{F}_t = \sigma\big(W(t)\big)$ is the sigma-algebra generated by $W(t)$. Within this framework, where we assume the flow of information is given only by the observed values of W, it is clear that any decision taken at time t should take into account only values of the process W up to time t. It is also required that two actions can not take place at the same time, i.e., we do not allow two stopping times to occur at the same discrete time instant. These assumptions are precisely stated in the following definition, and for further details on the theory of multiple optimal stopping rules we refer the reader to [Nikolaev and Sofronov, 2007] and [Sofronov, 2013].

Definition 7.3.1. *A collection of integer-valued random variables $(\tau_1, \ldots, \tau_i)$ is called an i-multiple stopping rule if the following conditions hold:*

(a) $\{\omega \in \Omega \ : \ \tau_1(\omega) = m_1, \ldots, \tau_j(\omega) = m_j\} \in \mathcal{F}_{m_j}, \ \forall m_j > m_{j-1} > \ldots > m_1 \geq 1, \ j = 1, \ldots, i;$

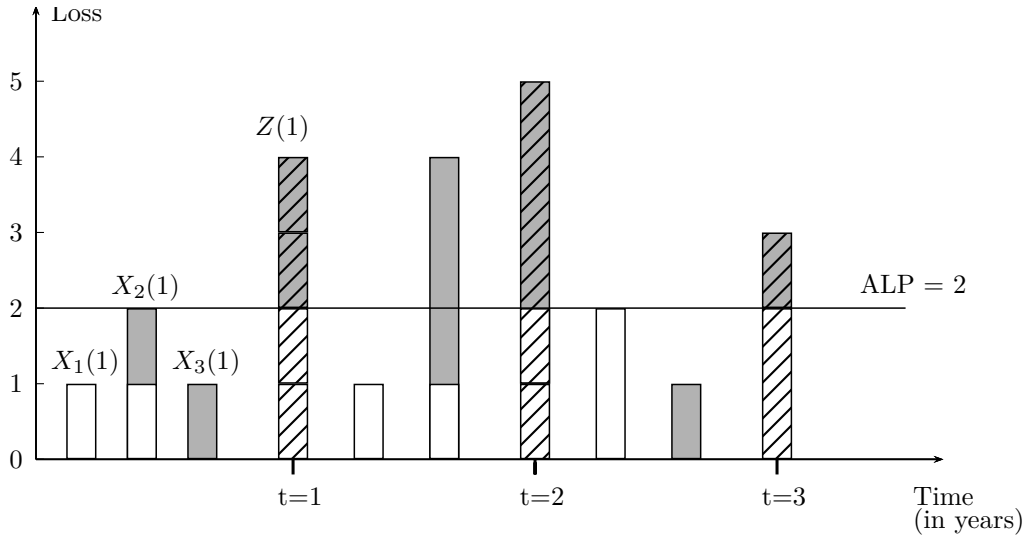

Figure 7.3: Accumulated Loss Policy (ALP) with ALP level of 2.0.

(b) $1 \leq \tau_1 < \tau_2 < \ldots < \tau_i < +\infty$, ($\mathbb{P}$-*a.s.*).

Given the mathematical definition of a stopping rule the notion of optimality of these rules can be made precise in the following definitions.

Definition 7.3.2. *For a given multiple stopping rule* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)$ *the gain function utilized in this chapter takes the following additive form:*

$$g(\boldsymbol{\tau}) = W(\tau_1) + \ldots + W(\tau_k).$$

Definition 7.3.3. *Let* $\mathcal{S}_m$ *be the class of multiple stopping rules* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)$ *such that* $\tau_1 \geq m$ ($\mathbb{P}$-*a.s.*). *The function*

$$v_m = \sup_{\boldsymbol{\tau} \in \mathcal{S}_m} \mathbb{E}[g(\boldsymbol{\tau})],$$

is defined as the m-value of the game and, in particular, if $m = 1$ *then* v_1 *is the value of the game.*

Definition 7.3.4. *A multiple stopping rule* $\boldsymbol{\tau}^* \in \mathcal{S}_m$ *is called an optimal multiple stopping rule in* $\mathcal{S}_m$ *if* $\mathbb{E}[W(\boldsymbol{\tau}^*)]$ *exists and* $\mathbb{E}[W(\boldsymbol{\tau}^*)] = v_m$.

The following result (first presented in [Nikolaev and Sofronov, 2007], Theorem 3) provides the optimal multiple stopping rule that maximizes the expectation of the sum of the observations (see Figure 7.3 for a schematic representation).

Theorem 7.3.5. *Let* $W(1), W(2), \ldots, W(T)$ *be a sequence of independent random variables with known distribution functions* $F_1, F_2, \ldots, F_T$, *and the gain function* $g(\boldsymbol{\tau}) = \sum_{j=1}^{k} W(\tau_j)$. *Let* $v^{L,l}$ *be the value of a game where the agent is allowed to stop* l *times* ($l \leqslant k$) *and there are* L ($L \leqslant T$) *steps remaining. If there exist* $\mathbb{E}[W(1)], \mathbb{E}[W(2)], \ldots, \mathbb{E}[W(T)]$ *then the value of the*

game is given by

$$\begin{aligned} v^{1,1} &= \mathbb{E}[W(T)], \\ v^{L,1} &= \mathbb{E}\big[\max\{W(T-L+1), v^{L-1,1}\}\big], \quad 1 < L \leq T, \\ v^{L,l+1} &= \mathbb{E}\big[\max\{v^{L-1,l} + W(T-L+1), v^{L-1,l+1}\}\big], \quad l+1 < L \leq T, \\ v^{l,l} &= \mathbb{E}\left[v^{l-1,l-1} + W(T-l+1)\right]. \end{aligned}$$

If we put

$$\begin{aligned} &\tau_1^* = \min\{m_1 \,:\, 1 \leqslant m_1 \leqslant T-k+1, W(m_1) \geqslant v^{T-m_1,k} - v^{T-m_1,k-1}\}; \\ &\tau_i^* = \min\{m_i \,:\, \tau_{i-1}^* < m_i \leqslant T-k+i, W(m_i) \geqslant v^{T-m_i,k-i+1} - v^{T-m_i,k-i}\}, \ i = 2, \ldots, k-1; \\ &\tau_k^* = \min\{m_k \,:\, \tau_{k-1}^* < m_k \leqslant T, W(m_k) \geqslant v^{T-m_k,1}\}; \end{aligned} \tag{7.1}$$

then $\boldsymbol{\tau}^* = (\tau_1^*, \ldots, \tau_k^*)$ *is the optimal multiple stopping rule.*

In the context we consider it is always be optimal to stop the process exactly k times, but this may not be true, for example, if some reward is given to the product holder for less than k years of claims of insurance. In the absence of such considerations, we proceed with assuming always k years of claims will be made. In Theorem 7.3.5 we see that the value function for $L > l$ is artificial and $v^{0,1}$, for example, has no interpretation. On the other hand, $v^{1,1}$ can not be calculated using the general formula (it would depend on $v^{0,1}$). With one stop remaining and one step left, from the reasons given above, we are obliged to stop, and, therefore, there is no maximization step when calculating $v^{1,1}$, i.e., $v^{1,1} = \mathbb{E}[W(T-1+1)]$. The same argument is valid for $l > 1$ and, in this case,

$$v^{L,l} = \mathbb{E}\left[\max\{v^{L-1,l-1} + W(T-L+1),\ v^{L-1,l}\}\right],\ 1 \leq l \leq T,$$

and, if we have $l \leq (T-1)$ steps left and also l stops, we must stop in all the steps remaining. So,

$$v^{l,l} = \mathbb{E}\left[v^{l-1,l-1} + W(T-l+1)\right].$$

From Theorem 7.3.5 and the assumption of independence of the annual losses, we see that to be able to calculate the optimal rule we only need to calculate (unconditional) expectations of the form $\mathbb{E}[W]$ and $\mathbb{E}[\max\{c_1 + W,\ c_2\}]$, for different values of c_1 and c_2. In addition, since $0 \leq v^{L-1,l} \leq v^{L-1,l+1}$, we actually only need to calculate $\mathbb{E}[\max\{c_1 + W,\ c_2\}]$ for $0 \leq c_1 \leq c_2$.

7.3.1 Objective functions for rational and boundedly rational insurees

In this section we consider two possible general populations for the potential insuree. The first group are those that are perfectly rational, meaning that they always act in an optimal fashion when given the chance and, more importantly, are capable (i.e. have the resources) to figure out what is the optimal behaviour. In this case we consider a global objective function to be optimized.

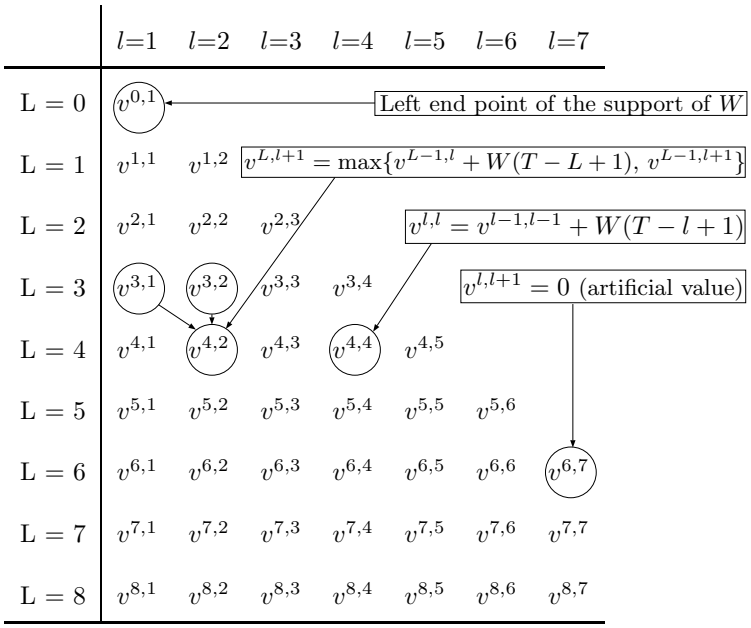

Figure 7.4: Schematic representation of the value function iteration.

The second group represent boundedly rational insurees who act sub-optimally. This group represents firms who are incapable or lack the resources/knowledge to understand how to act optimally when determining their optimal behaviours/actions and are captured by local behaviours.

Hence, these two populations are encoded in two objective functions: one which is optimal (globally) and one which represents a sub-optimal (local strategy) the boundedly rational population would likely adopt. These behaviours can be made precise through the following exercising strategies, for the first and second groups, respectively.

1. **Global Risk Transfer Strategy:** Minimizes the (expected) total loss over the period $[0, T]$;

2. **Local Risk Transfer Strategy:** Minimizes the (expected) sum of the losses at the insurance times (i.e. stopping times).

These two different groups can be understood as, for example, large corporations, with employees dedicated to fully understand the mathematical nuances of this kind of contract and small companies, with limited access to information. The group with "bounded rationality" may decide (heuristically, without the usage of any mathematical tool) to follow the so-called Local Risk Transfer Strategy, which produces smaller gain in the period $[0, T]$. As we will see in Section 7.6 these two different objective functions can lead to completely different exercising strategies, and we believe the insurance company who sells this contract should be aware of these different behaviours.

For the first loss function the formal objective is to minimize

$$\sum_{\substack{t=1\\ t\notin\{\tau_1,\ldots,\tau_k\}}}^{T} Z(t) + \sum_{j=1}^{k} \widetilde{Z}(\tau_j) = \sum_{t=1}^{T} Z(t) - \sum_{\substack{t=1\\ t\in\{\tau_1,\ldots,\tau_k\}}}^{T} \Big\{ Z(t) - \widetilde{Z}(t) \Big\}.$$

Since $\sum_{t=1}^{T} Z(t)$ does not depend on the choice of $\tau_1, \ldots, \tau_k$, this is, in fact, equivalent to maximize

$$\sum_{j=1}^{k} W(\tau_j) = \sum_{j=1}^{k} \Big\{ Z(\tau_j) - \widetilde{Z}(\tau_j) \Big\},$$

where the process W is defined as $W(t) = Z(t) - \widetilde{Z}(t)$.

For the second objective function, the company aims to minimise the total loss not over period $[0, T]$ but instead only at times at which the decisions are taken to apply insurance and therefore claim against losses in the given year,

$$\sum_{j=1}^{k} \widetilde{Z}(\tau_j),$$

and, in this case, the process W should be viewed as $W(t) = -\widetilde{Z}(t)$.

Remark 7.3.6. *Note that if the agent is trying to maximize the first loss function (using $W = Z - \widetilde{Z}$), then W is non-negative stochastic process, and only one kind of expectation is required to be calculated, since if $c_1 = c_2 = 0$, then $\mathbb{E}[\max\{c_1 + W,\ c_2\}] = \mathbb{E}[W]$.*

Remark 7.3.7. *If the agent is trying to minimize the expected gain of the sum of $\widetilde{Z}(t)$ random variables (instead of maximizing it) one can rewrite the problem as follows. Define a process $W(t) = -\widetilde{Z}(t)$ and note that $\min \mathbb{E}[\sum_{j=1}^{k} \widetilde{Z}(\tau_j)] = \max \mathbb{E}[\sum_{j=1}^{k} W(\tau_j)]$. Therefore the optimal stopping times that maximize the expected sum of the process W are the same that minimize the expected sum of the process $\widetilde{Z}$.*

Although this work is mainly devoted to study of a combination of insurance policy and severity distribution that leads to closed form results of the value function integrals required for closed form multiple optimal stopping rules, we also show how one can develop principled approximation procedures can be used to calculate the distribution of the insured process $\widetilde{Z}$ (see Section 7.7). In the remainder of this section we present a very simple example using the second (local) objective function, where we assume the annual insured losses are modelled as Log-Normal random variables.

Example 7.3.8 (Log-Normal)**.** *To give us some intuition, let us assume that the insured losses $\widetilde{Z}(1), \ldots, \widetilde{Z}(T)$ form a sequence of i.i.d. random variables such that $\widetilde{Z} \sim$ Log-Normal(0,1). Note that this would not be a reasonable assumption in practice, as we know the true distribution of the insured losses has a point mass at zero. To calculate the multiple optimal rule that minimizes the expected loss let us define $W = -\widetilde{Z}$. The values of the game using the equations in Theorem 7.3.5 can be seen in Table 7.1.*

Note that Table 7.1 presents the value of expected loss at the times we stop*, i.e.* , $\mathbb{E}\left[\sum_{j=1}^{k} -\widetilde{Z}(\tau_j)\right]$, *so it only makes sense to compare values within the same column. Doing so one can see that for a fixed number of stops l, the value of the game is increasing with the number of steps remaining. In other words the more one can wait to decide in which step to stop the smaller is the expected loss.*

L	l=1	l=2	l=3	l=4	l=5	l=6	l=7	l=8	l=9
0	0.00								
1	-1.65	0.00							
2	-1.02	-3.30	0.00						
3	-0.77	-2.19	-4.95	0.00					
4	-0.64	-1.71	-3.45	-6.59	0.00				
5	-0.55	-1.43	-2.76	-4.77	-8.24	0.00			
6	-0.49	-1.25	-2.34	-3.87	-6.12	-9.89	0.00		
7	-0.44	-1.12	-2.05	-3.32	-5.04	-7.51	-11.54	0.00	
8	-0.41	-1.02	-1.85	-2.94	-4.36	-6.26	-8.91	-13.19	0.00
9	-0.38	-0.94	-1.69	-2.65	-3.88	-5.45	-7.50	-10.34	-14.84
10	-0.36	-0.88	-1.56	-2.43	-3.52	-4.87	-6.58	-8.78	-11.78

Table 7.1: Table of the value function for different L (steps remaining) and l stops in the Log-Normal example.

If we suppose that $T = 7$, and we are granted four stops the expected loss is $v^{7,4} = -3.32$. In this case the optimal stopping rule is given by

$$\begin{aligned}
\tau_1^* &= \min\{m_1 \,:\, 1 \leqslant m_1 \leqslant 4, W(m_1) \geqslant v^{7-m_1,4} - v^{7-m_1,3}\},\\
\tau_2^* &= \min\{m_2 \,:\, \tau_1^* < m_2 \leqslant 5, W(m_2) \geqslant v^{7-m_2,3} - v^{7-m_2,2}\},\\
\tau_3^* &= \min\{m_3 \,:\, \tau_2^* < m_3 \leqslant 6, W(m_3) \geqslant v^{7-m_3,2} - v^{7-m_3,1}\},\\
\tau_4^* &= \min\{m_4 \,:\, \tau_3^* < m_4 \leqslant 7, W(m_4) \geqslant v^{7-m_4,1}\}.
\end{aligned}$$

For instance, if we observe the sequence

$$w_1 = -0.57, w_2 = -0.79, w_3 = -4.75, w_4 = -1.07, w_5 = -1.14, w_6 = -5.56, w_7 = -1.59,$$

then the optimal stopping times are given by:

$$\begin{aligned}
&\tau_1^* = 1, \ \textit{since } w_1 = -0.57 \geq v^{7-1,4} - v^{7-1,3} = -3.87 - (-2.34) = -1.53;\\
&\tau_2^* = 2, \ \textit{since } w_2 = -0.79 \geq v^{7-2,3} - v^{7-2,2} = -2.76 - (-1.43) = -1.33;\\
&\tau_3^* = 4, \ \textit{since } w_4 = -1.07 \geq v^{7-4,2} - v^{7-4,1} = -2.19 - (-0.77) = -1.42;\\
&\tau_4^* = 7, \ \textit{because we are obliged to stop exactly 4 times.}
\end{aligned}$$

In this case the realized loss at the stopping times *is $-0.57 - 0.79 - 1.07 - 1.59 = -4.02$, wich should be comparable with the expected loss under the optimal rule: -3.32.*

7.4 Loss process models via Loss Distributional Approach

Before discussing the application of the Theorem 7.3.5 to the problem of choosing the multiple exercising dates of the insurance product present in Section 7.1, in this section we present the LDA model that leads to closed form solutions in Section 7.5.

The Loss Distributional Approach (LDA) in OpRisk assumes that during a year t a company suffers $N(t)$ operational losses, with $N(t)$ following some counting distribution (usually Poisson or Negative Binomial). The severity of each of these losses is denoted by $X_1(t), \ldots, X_{N(t)}(t)$ and the cumulative loss by the end of year t is given by $Z(t) = \sum_{n=1}^{N(t)} X_n(t)$. For the purpose of modelling OpRisk losses it is essential that the severity density allows extreme events to occur, since these events often occur in practice, as shown, for example, in [Peters et al., 2013, Section 1.1]. Following the nomenclature in [Franzetti, 2011, Table 3.3], the Inverse Gaussian distribution possess a "moderate tail" which makes it a reasonable model for OpRisk losses for many risk process types and is often used in practice. This family of distributions also has the advantage of being closed under convolution and this characteristic is essential if closed form solutions for the multiple optimal stopping problem are to be obtained.

In the closed form solutions we present for the different insurance policies we use properties of the Inverse Gaussian distribution and its relationship with the Generalized Inverse Gaussian distribution. The following Lemmas will be used throughout; see additional details in [Folks and Chhikara, 1978] and [Jørgensen, 1982].

In the following, let $X_1, \ldots, X_n$ be a sequence of i.i.d. Inverse Gaussian (IG) random variable with parameters $\mu, \lambda > 0$, ie,

$$f_X(x;\ \mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2} x^{-3/2} \exp\left\{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

Let also G be a Generalized Inverse Gaussian (GIG) r.v. with parameters $\alpha, \beta > 0$, $p \in \mathbb{R}$, i.e.,

$$f_G(x;\ \alpha, \beta, p) = \frac{(\alpha/\beta)^{p/2}}{2K_p(\sqrt{\alpha\beta})} x^{p-1} \exp\left\{-\frac{1}{2}(\alpha x + \beta/x)\right\}, \quad x > 0,$$

where K_p is a modified Bessel function of the third kind (sometimes called modified Bessel function of the second kind), defined as

$$K_p(z) = \frac{1}{2}\int_0^{+\infty} u^{p-1} e^{-z(u+1/u)/2} du.$$

Lemma 7.4.1. *The Inverse Gaussian family of random variables is closed under convolution and the distribution of its sum is given by*

$$S_n := \sum_{l=1}^{n} X_l \sim IG(n\mu,\ n^2\lambda). \tag{7.2}$$

Lemma 7.4.2. *Any Inverse Gaussian random variable can be represented as a Generalized Inverse Gaussian, and for the particular case of Lemma 7.4.1 the relationship is*

$$f_{S_n}(x;\ n\mu, n^2\lambda) \equiv f_G(x;\ \lambda/\mu^2, n^2\lambda, -1/2). \tag{7.3}$$

Lemma 7.4.3. *Modified Bessel functions of the third kind are symmetric around zero in the parameter p. In particular when $p = 1/2$,*

$$\frac{K_{1/2}(\frac{n\lambda}{\mu})}{K_{-1/2}(\frac{n\lambda}{\mu})} = 1. \tag{7.4}$$

Lemma 7.4.4. *The density of an Inverse Gaussian r.v. has the following property (which clearly holds for any power of x, with the proper adjustment in the last parameter of the GIG in the right hand side):*

$$x f_G(x;\ \lambda/\mu^2, n^2\lambda, -1/2) \equiv n\mu\ f_G(x;\ \lambda/\mu^2, n^2\lambda, 1/2). \tag{7.5}$$

Proof. (of Lemmas 7.4.1–7.4.4) The proof of Lemma 7.4.1 can be found in [Tweedie, 1957, Section 2] and the result in Lemma 7.4.2 can be seen by comparing the kernel of both distributions.

The symmetry in Lemma 7.4.3 can be seen through the following characterization of modified Bessel functions of the third kind

$$K_p(x) := \int_0^{+\infty} \exp\{-x\cosh(t)\}\cosh(pt)dt,$$

(see [Watson, 1922], page 181) and the fact that $\cosh(-p) = (-1)\cosh(p)$. The last result, Lemma 7.4.4, follows from Lemma 7.4.3 and a simple comparison of the densities. □

7.5 Closed-form multiple optimal stopping rules for multiple insurance usage decisions

In this section we present some models in which the optimal rules can be calculated explicitly, with all the technical proofs postponed to the Appendix. Using the results presented in Section 7.4 we show that if we assume a Poisson-Inverse Gaussian LDA model, where $X_n \sim IG(\lambda, \mu)$ and $N \sim Poi(\lambda_N)$, the optimal times (years) to exercise or make claims on the insurance policy for the Accumulated Loss Policy (ALP) can be calculated analytically regardless of where the global or local gain (objective) functions are considered. For the Individual Loss Policy (ILP), when using the local objective function we propose to model the losses after the insurance policy is applied and, in this case, we present analytical solutions for the stopping rules. On the other hand, the ILP under the total loss case given by the global objective function does not produce a closed form solution. However, we show how a simple Monte Carlo scheme can be used to accurately estimate the results.

Since we assume the annual losses $Z(1), \ldots, Z(T)$ are identically distributed we denote by Z a r.v. such that $Z \sim Z(1)$. As in the previous sections, $\widetilde{Z}$ is the insured process; $S_n = \sum_{k=1}^{n} X_k$ is the partial sum up to the n-th loss and $p_m = \mathbb{P}[N = m]$ is the probability of observing m losses in one year. The gain W is defined as either $-\widetilde{Z}$, when the objective is to minimize the loss at the times the company uses the insurance policy (local optimality), or $Z - \widetilde{Z}$, in case the function to be minimized is the total loss over the time horizon $[0, T]$, i.e. (global optimality).

7.5.1 Accumulated Loss Policy (ALP)

For the ALP case (see Definition 7.2.2) we can model the severity of the losses before applying the insurance policy. Conditional upon the fact that $\sum_{n=1}^{m} X_n > ALP$, then the annual loss after the application of the insurance policy is $\sum_{n=1}^{m} X_n - ALP$. With this in mind, we calculate the c.d.f.'s of the insured process, $\widetilde{Z}$ and also of the random variable $Z - \widetilde{Z}$.

7.5.1.1 Local risk transfer objective: Minimizing the loss at the stopping times

Proposition 7.5.1 (Local Risk Transfer Case)**.** *The c.d.f. and p.d.f. of the insured process are given, respectively, by*

$$F_{\widetilde{Z}}(z) = \sum_{m=1}^{+\infty} F_{IG}(z + ALP;\ m\mu, m^2\lambda)p_m + C_0, \tag{7.6}$$

$$f_{\widetilde{Z}}(z) = \sum_{m=1}^{+\infty} \Big\{ f_{IG}(z + ALP;\ m\mu, m^2\lambda)p_m \Big\} \mathbb{1}_{\{z>0\}} + C_0 \mathbb{1}_{\{z=0\}}; \tag{7.7}$$

where the constant C_0 is defined as $C_0 := \sum_{m=1}^{+\infty} F_{IG}(ALP;\ m\mu, m^2\lambda)p_m + p_0$.

Proof.

$$\begin{aligned}
F_{\widetilde{Z}}(z) &= \mathbb{P}[\widetilde{Z} \le z] \\
&= \mathbb{P}[\max\{Z - ALP,\ 0\} \le z] \\
&= \mathbb{P}[\max\{Z - ALP,\ 0\} \le z \mid Z - ALP > 0]\ \mathbb{P}[Z - ALP > 0] \\
&\quad + \mathbb{P}[\max\{Z - ALP,\ 0\} \le z \mid Z - ALP \le 0]\ \mathbb{P}[Z - ALP \le 0] \\
&= \mathbb{P}[\max\{Z - ALP,\ 0\} \le z \mid Z > ALP]\ \mathbb{P}[Z > ALP] \\
&\quad + \mathbb{P}[0 \le z \mid Z \le ALP]\ \mathbb{P}[Z \le ALP] \\
&= \mathbb{P}[A < Z \le z + A] + \mathbb{1}_{\{z\ge 0\}}(z) F_Z(A) \\
&= \big[F_Z(z + A) - F_Z(A)\big]\mathbb{1}_{\{z>0\}}(z) + \mathbb{1}_{\{z>0\}}(z) F_Z(A) + \mathbb{1}_{\{z=0\}}(z) F_Z(A) \\
&= F_Z(z + A)\mathbb{1}_{\{z>0\}}(z) + F_Z(A)\mathbb{1}_{\{z=0\}}(z) \\
&= F_Z(z + A)\mathbb{1}_{\{z\ge 0\}}(z) \\
&= \sum_{m=1}^{\infty} p_m F_{S_m}(z + A) + p_0 \\
&= \sum_{m=1}^{\infty} p_m F_{IG}(z + A;\ m\mu,\ m^2\lambda) + p_0.
\end{aligned}$$

The p.d.f. easily follows from the derivation of $F_{\widetilde{Z}}(z)$ with respect to z but it is important to note that $f_{\widetilde{Z}}$ is a continuous density with discrete mass at $z = 0$, ie,

$$f_{\widetilde{Z}}(z) = \sum_{m=1}^{+\infty} \Big\{ f_{IG}(z + ALP;\ m\mu, m^2\lambda)p_m \Big\} \mathbb{1}_{\{z>0\}} + \underbrace{\Big\{ p_0 + \sum_{m=1}^{+\infty} F_{IG}(ALP;\ m\mu, m^2\lambda)p_m \Big\}}_{\mathbb{P}[\widetilde{Z}=0]} \mathbb{1}_{\{z=0\}}.$$

□

After calculating the distribution of $\widetilde{Z}$ we can calculate expectations of the form $\mathbb{E}\left[\max\left\{c_1+W, c_2\right\}\right]$ w.r.t. the loss process Z and, therefore one can consequently obtain the multiple optimal stopping rules under the Accumulated Loss Policy via direct application of Theorem 7.3.5.

Theorem 7.5.2 (Local Risk Transfer Case)**.** *Using the notation of Theorem 7.3.5 and defining $W(t) = -\widetilde{Z}(t)$, for $t = 1, \ldots, T$ the multiple optimal stopping rule is given by the set of equations in (7.1), where*

$$
\begin{aligned}
v^{1,1} &= -\sum_{m=1}^{+\infty} p_m \Big(m\mu \overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - ALP\overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Big), \\
v^{L,1} &= -\sum_{m=1}^{+\infty} p_m \Bigg[\Bigg(m\mu \Big(F_{GIG}(v^{L-1,1} + ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) \Big) \\
&\quad + ALP \Big(F_{GIG}(v^{L-1,1} + ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Big) \Bigg) \\
&\quad + v^{L-1,1} \overline{F}_{GIG}(v^{L-1,1} + ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Bigg], \\
v^{L,l+1} &= -\sum_{m=1}^{+\infty} p_m \Bigg[\Bigg(m\mu \Big(F_{GIG}(v^{L-1,l+1} - v^{L-1,l} + ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) \Big) \\
&\quad - (v^{L-1,l} - ALP) \Big(F_{GIG}(v^{L-1,l+1} - v^{L-1,l} + ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Big) \Bigg) \\
&\quad + v^{L-1,l+1} \overline{F}_{GIG}(v^{L-1,l+1} - v^{L-1,l} + ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Bigg] - v^{L-1,l} C_0, \\
v^{l,l} &= v^{l-1,l-1} - \sum_{m=1}^{+\infty} p_m \Big(m\mu \overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - ALP\overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Big).
\end{aligned}
$$

Proof. As in Theorem 7.5.5, to calculate the optimal rule we only need to calculate $\mathbb{E}[W]$ and $\mathbb{E}[\max\{-c_1+W,\ -c_2\}]$, for $0 < c_1 < c_2$. Given the expression (7.7) for the density of $\widetilde{Z}$ we can calculate $\mathbb{E}[W]$ as follows

$$
\begin{aligned}
\mathbb{E}[\widetilde{Z}] &= \int_0^{+\infty} z \sum_{m=1}^{+\infty} f_{IG}(z + ALP;\ m\mu, m^2\lambda) p_m dz \\
&= \sum_{m=1}^{+\infty} p_m \int_0^{+\infty} z f_{GIG}(z + ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) dz \\
&= \sum_{m=1}^{+\infty} p_m \int_{ALP}^{+\infty} (w - ALP) f_{GIG}(w;\ \lambda/\mu^2, m^2\lambda, -1/2)\ dw \\
&= \sum_{m=1}^{+\infty} p_m \Big(m\mu \overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - ALP\overline{F}_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \Big)
\end{aligned}
$$

And then we use the fact that $\mathbb{E}[W] = -\mathbb{E}[\widetilde{Z}]$.

For the second term we have that $\mathbb{E}[\max\{-c_1+W,\ -c_2\}] = (-1)\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}]$ and

$$\begin{aligned}
\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}] &= \int_0^{+\infty} \min\{c_1+z,\ c_2\} f_{\widetilde{Z}}(z)dz \\
&= \sum_{m=1}^{+\infty} p_m \int_0^{+\infty} \min\{c_1+z,\ c_2\} f_{IG}(z+ALP;\ m\mu, m^2\lambda)dz \\
&\quad + \min\{c_1+0,\ c_2\}\Big\{p_0 + \sum_{m=1}^{+\infty} F_{IG}(ALP;\ m\mu, m^2\lambda)p_m\Big\} \\
&= \sum_{m=1}^{+\infty} p_m \left[\int_{ALP}^{+\infty} \min\{c_1+w-ALP,\ c_2\} f_{IG}(w;\ m\mu, m^2\lambda)dw\right] + c_1C_0 \\
&= \sum_{m=1}^{+\infty} p_m \left[\int_{ALP}^{c_2-c_1+ALP} (c_1+w-ALP) f_{IG}(w;\ m\mu, m^2\lambda)dw\right. \\
&\quad \left. + \int_{c_2-c_1+ALP}^{+\infty} c_2 f_{IG}(w;\ m\mu, m^2\lambda)dw\right] + c_1C_0 \\
&= \sum_{m=1}^{+\infty} p_m \left[\left(m\mu\Big(F_{GIG}(c_2-c_1+ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2)\Big)\right.\right. \\
&\quad + (c_1-ALP)\Big(F_{GIG}(c_2-c_1+ALP;\ \lambda/\mu^2, m^2\lambda, -1/2) \\
&\quad \left.\left. - F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, -1/2)\Big)\right) + c_2\overline{F}_{GIG}(c_2-c_1+ALP;\ \lambda/\mu^2, m^2\lambda, -1/2)\right] \\
&\quad + c_1C_0.
\end{aligned}$$

□

7.5.1.2 Global risk transfer objective: Minimizing the loss over the period $[0,T]$

If we assume the company wants to minimize its total loss over the period $[0,T]$ the gain achieved through the Accumulated Loss Policy (ALP) is given by

$$\begin{aligned}
W &= Z - \widetilde{Z} \\
&= \sum_{n=1}^{N} X_n - \left(\sum_{n=1}^{N} X_n - ALP\right) \mathbb{1}_{\left\{\sum_{n=1}^{N} X_n > ALP\right\}} \\
&= ALP\mathbb{1}_{\left\{\sum_{n=1}^{N} X_n > ALP\right\}} + \left(\sum_{n=1}^{N} X_n\right) \mathbb{1}_{\left\{\sum_{n=1}^{N} X_n > ALP\right\}} \\
&= \min\left\{ALP,\ \sum_{n=1}^{N} X_n\right\}.
\end{aligned}$$

For notational convenience we denote by $W_m = \min\{ALP,\ \sum_{n=1}^{m} X_n\}$ the annual gain conditional on the fact that m losses were observed.

Proposition 7.5.3 (Global Risk Transfer Case: ALP)**.** *The c.d.f. and p.d.f. of the gain process are given, respectively, by*

$$F_W(w) = \mathbb{1}_{\{w\geq ALP\}} + F_{S_m}(w)\mathbb{1}_{\{w<ALP\}}, \tag{7.8}$$

$$f_W(w) = \sum_{m=1}^{N} \Big\{\Big(\overline{F}_{S_m}(ALP)\mathbb{1}_{\{w=ALP\}} + f_{S_m}(w)\mathbb{1}_{\{0<w<ALP\}}\Big)p_m\Big\} + p_0\mathbb{1}_{\{w=0\}}. \tag{7.9}$$

Proof. For $w \geq ALP$, we clearly have $F_W(w) = \mathbb{P}[\min\{ALP,\ Z\} \leq w] = 1$. For $0 \leq w < ALP$,

$$\begin{aligned}
F_W(w) &= \mathbb{P}[W \leq w] \\
&= \mathbb{P}[\min\{ALP,\ Z\} \leq w] \\
&= \mathbb{P}[\min\{ALP,\ Z\} \leq w \mid Z \leq ALP]\ \mathbb{P}[Z \leq ALP] \\
&\quad + \mathbb{P}[\min\{ALP,\ Z\} \leq w \mid Z > ALP]\ \mathbb{P}[Z > ALP] \\
&= \mathbb{P}[Z \leq w \mid Z \leq ALP]\ \mathbb{P}[Z \leq ALP] + \mathbb{P}[ALP \leq w \mid Z > ALP]\ \mathbb{P}[Z > ALP] \\
&= \mathbb{P}[Z \leq w,\ Z \leq ALP] + \mathbb{1}_{\{w \geq ALP\}}(w)\overline{F}_Z(ALP) \\
&= \mathbb{P}[Z \leq \min\{w,\ ALP\}] + \mathbb{1}_{\{w \geq ALP\}}(w)\overline{F}_Z(ALP).
\end{aligned}$$

Since we assumed $w < ALP$, we have that $\min\{w,\ ALP\} = w$ and the indicator function on the second term is always equal to zero leading to the following expression for an arbitrary $w \geq$:

$$\begin{aligned}
F_W(w) &= \mathbb{1}_{\{0 \leq w < ALP\}}(w)F_Z(w) + \mathbb{1}_{\{w \geq ALP\}}(w) \\
&= \mathbb{1}_{\{0 \leq w < ALP\}}(w)\Big\{p_0 + \sum_{m=1}^{\infty} F_{S_m}(w)p_m\Big\} + \mathbb{1}_{\{w \geq ALP\}}(w) \\
&= \mathbb{1}_{\{0 \leq w < ALP\}}(w)\Big\{p_0 + \sum_{m=1}^{\infty} F_{IG}(w;\ m\mu,\ m^2\lambda)p_m\Big\} + \mathbb{1}_{\{w \geq ALP\}}(w).
\end{aligned}$$

Consequently, the pdf of the gain is given by

$$f_W(w) = \sum_{m=1}^{N} \Big\{\Big(\overline{F}_{S_m}(ALP)\mathbb{1}_{\{w=ALP\}} + f_{S_m}(w)\mathbb{1}_{\{0<w<ALP\}}\Big)p_m\Big\} + p_0\mathbb{1}_{\{w=0\}}.$$

□

After calculating the distribution of the gain, W, we can calculate expectations w.r.t. it and, therefore, the multiple optimal stopping rule under the Accumulated Loss Policy is obtained via direct application of Theorem 7.3.5.

Theorem 7.5.4 (Global Risk Transfer Case: ALP)**.** *Defining $W(t) = Z(t) - \widetilde{Z}(t)$, for $t = 1, \ldots, T$ the multiple optimal stopping rule is given by (7.1), where*

$$\begin{aligned}
v^{1,1} &= \sum_{m=1}^{+\infty} p_m\Big\{\overline{F}_{S_m}(ALP)ALP + m\mu F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2)\Big\}, \\
v^{L,1} &= \sum_{m=1}^{+\infty} p_m\Big\{\overline{F}_{S_m}(ALP)\max\{ALP,\ v^{L-1,1}\} + m\mu\big(F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) \\
&\quad - F_{GIG}(v^{L-1,1};\ \lambda/\mu^2, m^2\lambda, 1/2)\big) + v^{L-1,1}F_{S_m}(\min\{v^{L-1,1},\ ALP\})\Big\} + p_0v^{L-1,1}, \\
v^{L,l+1} &= \sum_{m=1}^{+\infty} p_m\Big\{\overline{F}_{S_m}(ALP)\max\{v^{L-1,l} + ALP,\ v^{L-1,l+1}\} \\
&\quad + v^{L-1,l}(F_{S_m}(ALP) - F_{S_m}(v^{L-1,l+1} - v^{L-1,l})) + m\mu\big(F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) \\
&\quad - F_{GIG}(v^{L-1,l+1} - v^{L-1,l};\ \lambda/\mu^2, m^2\lambda, 1/2)\big) + v^{L-1,l+1}F_{S_m}(\min\{v^{L-1,l+1} - v^{L-1,l},\ ALP\})\Big\} \\
&\quad + p_0v^{L-1,l+1}, \\
v^{l,l} &= \sum_{m=1}^{+\infty} p_m\Big\{\overline{F}_{S_m}(ALP)ALP + m\mu F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2)\Big\}.
\end{aligned}$$

Proof. For $0 < c_1 < c_2$, the quantity of interest can be calculated as

$$\begin{aligned}
\mathbb{E}[\max\{c_1 + W, c_2\}] &= \int_0^{+\infty} \max\{c_1 + w, c_2\} f_W(w) dw \\
&= \sum_{m=1}^{+\infty} p_m \Big\{ \overline{F}_{S_m}(ALP) \max\{c_1 + ALP,\ c_2\} \\
&+ \int_{c_2 - c_1}^{ALP} (c_1 + w) f_{S_m}(w) dw + \int_0^{\min\{c_2 - c_1,\ ALP\}} c_2 f_{S_m}(w) dw \Big\} + p_0 c_2 \\
&= \sum_{m=1}^{+\infty} p_m \Big\{ \overline{F}_{S_m}(ALP) \max\{c_1 + ALP,\ c_2\} \\
&+ c_1 (F_{S_m}(ALP) - F_{S_m}(c_2 - c_1)) + m\mu \big(F_{GIG}(ALP;\ \lambda/\mu^2, m^2\lambda, 1/2) \\
&- F_{GIG}(c_2 - c_1;\ \lambda/\mu^2, m^2\lambda, 1/2) \big) + c_2 F_{S_m}(\min\{c_2 - c_1,\ ALP\}) \Big\} + p_0 c_2.
\end{aligned}$$

□

7.5.2 Individual Loss Policy (ILP)

The previous insurance policy, the ALP structure, has been based on the aggregated amount throughout the year. In the case of the ILP insurance structure, the coverage is not on an accumulated aggregate coverage, instead it is based on an individual loss event coverage.

7.5.2.1 Local risk transfer objective: minimizing the loss at the stopping times

Let us assume a company buys the insurance policy we call Individual Loss Policy (ILP). In this case, a particular loss process observed by the company *after* applying the insurance policy may be given by

$$X_1(\omega) - TCL,\ X_2(\omega) - TCL,\ 0,\ 0,\ 0,\ X_6(\omega) - TCL,\ 0, \ldots,\ X_{N-1}(\omega) - TCL,\ 0.$$

In this case we can define a new process $(\widetilde{X}_n)_{n \geq 1}$ such that

$$\widetilde{X}_1(\omega) := X_1(\omega) - TCL,\ \widetilde{X}_2(\omega) := X_2(\omega) - TCL,\ \widetilde{X}_3(\omega) := X_6(\omega) - TCL, \ldots, \widetilde{X}_{\widetilde{N}}(\omega) := X_{N-1}(\omega) - TCL$$

and the annual insured loss would be given by $\widetilde{Z} = \sum_{n=1}^{\widetilde{N}} \widetilde{X}_n$. Note that in this example the new process, $(\widetilde{X}_n)_{n \geq 1}$ would have $\widetilde{N} < N$ non zero observations and, in general, $\widetilde{N} \leq N$. The process $(\widetilde{X}_n)_{n \geq 1}$ can be interpreted as an auxiliary process, meaning that if the company had claimed on the insurance policy for this year then the observed losses would have been $\widetilde{X}_n$, instead of X_n.

In our approach we model the random variable $\widetilde{N}$ and the process $(\widetilde{X}_n)_{n \geq 1}$, the first as an homogeneous Poisson process with mean $\widetilde{\lambda}_{\widetilde{N}}$ and the second as a sequence of i.i.d. random variables such that $\widetilde{X} \sim IG(\lambda,\ \mu)$.

Theorem 7.5.5 (Local Risk Transfer Case: ILP)**.** *Assuming that $\widetilde{N} \sim Poi(\lambda_{\widetilde{N}})$ and $\widetilde{X}_1, \widetilde{X}_2, \ldots$ are i.i.d. with $\widetilde{X} \sim IG(\lambda,\ \mu)$ define $\widetilde{Z}(t) = \sum_{n=1}^{\widetilde{N}(t)} \widetilde{X}_n(t)$, and $W(t) = -\widetilde{Z}(t)$, for $t = 1, \ldots, T$.*

In this case the optimal stopping rule is given by (7.1), where

$$
\begin{aligned}
v^{1,1} &= -\lambda_{\widetilde{N}}\mu,\\
v^{L,1} &= -\sum_{n=1}^{+\infty}\Pr[\widetilde{N}=n]\Big[F_{GIG}(v^{L-1,1};\ \lambda/\mu^2, n^2\lambda, 1/2)n\mu\\
&\quad - v^{L-1,1}F_{GIG}(v^{L-1,1};\ \lambda/\mu^2, n^2\lambda, -1/2) + v^{L-1,1}\Big],\quad 1<L\leq T,\\
v^{L,l+1} &= -\sum_{n=1}^{+\infty}\Pr[\widetilde{N}=n]\Big[F_{GIG}(v^{L-l,l+1}-v^{L-l,l};\ \lambda/\mu^2, n^2\lambda, 1/2)n\mu\\
&\quad + (v^{L-l,l}-v^{L-l,l+1})F_{GIG}(v^{L-l,l+1}-v^{L-l,l};\ \lambda/\mu^2, n^2\lambda, -1/2) + v^{L-l,l+1}\Big]\\
&\quad + v^{L-l,l}\Pr[\widetilde{N}=0],\quad l+1<L\leq T,\\
v^{l,l} &= v^{l-1,l-1} - \lambda_{\widetilde{N}}\mu.
\end{aligned}
$$

Proof. It is clear from Theorem 7.3.5 that we only need to calculate two terms, namely $\mathbb{E}[W]$ and $\mathbb{E}[\max\{-c_1+W,\ -c_2\}]$, for $0<c_1<c_2$. The first term can be derived by a simple application of the Tower Property:

$$
\mathbb{E}[W] = -\mathbb{E}[\widetilde{Z}] = -\mathbb{E}\Big[\mathbb{E}[\widetilde{Z}|\widetilde{N}]\Big] = -\mathbb{E}[\widetilde{N}]\ \mathbb{E}[\widetilde{X}] = -\lambda_{\widetilde{N}}\mu.
$$

For the second term, first note that $\mathbb{E}[\max\{-c_1+W,\ -c_2\}] = (-1)\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}]$ and it then follows that, for $0<c_1<c_2$,

$$
\begin{aligned}
\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}] &= \int_0^{+\infty}\min\{c_1+z,\ c_2\}f_{\widetilde{Z}}(z)dz + \min\{c_1+0,\ c_2\}\Pr[\widetilde{N}=0]\\
&= \int_0^{+\infty}\big((c_1+z)\mathbb{1}_{\{c_1+z<c_2\}} + c_2\mathbb{1}_{\{c_1+z\geq c_2\}}\big)f_{\widetilde{Z}}(z)dz + c_1\Pr[\widetilde{N}=0]\\
&= \int_0^{c_2-c_1} z f_{\widetilde{Z}}(z)dz + c_1\int_0^{c_2-c_1} f_{\widetilde{Z}}(z)dz + c_2\int_{c_2-c_1}^{+\infty} f_{\widetilde{Z}}(z)dz + c_1\Pr[\widetilde{N}=0]\\
&= \sum_{n=1}^{+\infty}\Pr[\widetilde{N}=n]\Big[\int_0^{c_2-c_1} z f_{\widetilde{S}_n}(z)dz + c_1\int_0^{c_2-c_1} f_{\widetilde{S}_n}(z)dz + c_2\int_{c_2-c_1}^{+\infty} f_{\widetilde{S}_n}(z)dz\Big]\\
&\quad + c_1\Pr[\widetilde{N}=0]\\
&= \sum_{n=1}^{+\infty}\Pr[\widetilde{N}=n]\Big[F_{GIG}(c_2-c_1;\ \lambda/\mu^2, n^2\lambda, 1/2)n\mu + c_1F_{GIG}(c_2-c_1;\ \lambda/\mu^2, n^2\lambda, -1/2)\\
&\quad + c_2\overline{F}_{GIG}(c_2-c_1;\ \lambda/\mu^2, n^2\lambda, -1/2)\Big] + c_1\Pr[\widetilde{N}=0]\\
&= \sum_{n=1}^{+\infty}\Pr[\widetilde{N}=n]\Big[F_{GIG}(c_2-c_1;\ \lambda/\mu^2, n^2\lambda, 1/2)n\mu\\
&\quad + (c_1-c_2)F_{GIG}(c_2-c_1;\ \lambda/\mu^2, n^2\lambda, -1/2) + c_2\Big] + c_1\Pr[\widetilde{N}=0].
\end{aligned}
$$

Note that, for notational ease, $f_{\widetilde{Z}}$ must be understood as the absolutely continuous part of the density of $\widetilde{Z}$.

□

7.5.2.2 Global risk transfer objective: minimizing the loss over the period $[0, T]$ via Monte Carlo

If we assume the frequency of annual losses is given by $N \sim Poi(\lambda_N)$ and its severities by $X_i \sim IG(\lambda, \mu)$ then the gain process W is given by

$$\begin{aligned} W &= Z - \widetilde{Z} \\ &= \sum_{n=1}^{N} X_n - \sum_{n=1}^{N} \max\left(X_n - \text{TCL},\ 0\right) \\ &= \sum_{n=1}^{N} X_n - \sum_{n=1}^{N} \left(X_n - \text{TCL}\right) \mathbb{1}_{\{X_n > TCL\}} \\ &= \sum_{n=1}^{N} \left(TCL \mathbb{1}_{\{X_n > TCL\}} + X_n \mathbb{1}_{\{X_n \leq\}}\right) \\ &= \sum_{n=1}^{N} \min\{X_n,\ TCL\}. \end{aligned}$$

From Lemma 7.4.1 we know the Inverse Gaussian family is closed under convolution, but the distribution of the sum of truncated Inverse Gaussian r.v.'s does not take any known form. A simple and effective way to approximate the expectations necessary to the calculation of the optimal multiple stopping rule is to use a Monte Carlo scheme as follows.

Inputs: Model parameters $(\lambda, \mu, \lambda_N)$; Insurance policy parameter, TCL; Number of simulations M;

Result: Simple Random Sample from the gain r.v. W $(W^{(1)}, \ldots, W^{(M)})$;

for $i = 1, \ldots, M$ **do**

  Sample $N^{(i)} \sim Poi(\lambda_N)$;

  if $N = 0$ **then**

    $W^{(i)} = 0$;

  else

    Sample $X_k^{(i)} \sim IG(\lambda, \mu)$, for $k = 1, \ldots, N^{(i)}$;

    $W^{(i)} = \sum_{k=1}^{N^{(i)}} \min\{X_k^{(i)},\ TCL\}$

  end

end

By the end of this process we will have a sample $W^{(1)}, \ldots, W^{(M)}$ from the gain, which can be used to approximate, for any given values of $0 < c_1 < c_2$ the expectations as

$$\mathbb{E}[\max\{c_1 + W,\ c_2\}] \approx \frac{1}{M} \sum_{i=1}^{M} \max\{c_1 + W^{(i)},\ c_2\}.$$

7.6 Case studies

In this Section we analyse the results provided by the optimal rule in the scenario where analytical expressions are available. Although the loss distribution parameters are different for each

insurance policy, in this section we assume the insurance product is valid for $T = 8$ years and gives its owner the right to mitigate $k = 3$ losses.

First, for the Accumulated Loss Policy (ALP), Figure 7.5 presents a comparison of the two objective functions (Global and Local Risk Transfer), when the LDA parameters are $(\lambda, \mu, \lambda_N) = (3, 2, 3)$ and the insurance specific parameter is set to $ALP = 10$. In this case we know the probability of having an annual loss that would make it worth utilising the insurance product in one year is $\mathbb{P}[Z > ALP] \approx 20\%$. In this study, for a large number of scenarios, $M = 50,000$, the optimal rules from both objective functions were calculated as well as, for each scenario, the set of stopping times (m_1, m_2, m_3). On the bottom of Figure 7.5 we see that the exercising strategy is considerably different for the two objective functions. For the Global Risk Transfer, we see that fixing the first two stopping times, say $(m_1, m_2) = (1, 2)$, it is preferable (on average) to use the remaining right as early as possible. Another way to see the same pattern is to verify that the frequency of occurrence of the set of strategies $(1, 2, 3)$; $(2, 3, 4)$; $(3, 4, 5)$; $(4, 5, 6)$ is decreasing, again indicating a prevalence of early exercise strategies. On the other hand, if the objective is to minimize the "local risk", in more than 25% of the cases the optimal strategy will be to use the rights as soon as possible.

On the top of Figure 7.5 we present histograms of the total loss over $[0, T]$ (i) without insurance (solid line); (ii) using the global objective function (dark grey); (iii) using the local objective function (light grey). As expected the mean of the total loss when using the local loss function is greater than the global one, but still smaller than the total loss without any insurance.

For both the ALP and the ILP cases we check the optimality of the rules presented, comparing them with pre-specified stopping rules. Denoting (m_1, m_2, m_3) the three stopping times, the rules are defined as follows.

(i) Rule 1 (Deterministic): Always stops at $m_1 = 1, m_2 = 5, m_3 = 8$;

(ii) Rule 2 (Random): Stops randomly at three points in $(1, \ldots, 8)$, subject to $1 \leq m_1 < m_2 < m_3 \leq 8$;

(iii) Rule 3 (Average): Stops when the observed loss is less then the expected loss, i.e., $\mathbb{E}[W]$.

For a large number of scenarios, $M = 10,000$, we calculated the loss for each of the four rules (the Optimal, the Deterministic, the Random and the Average rules) and plot the histogram, comparing with the expected loss under the optimal rule, see Figure 7.6 for the Accumulated Loss Policy (ALP) and Figure 7.7 for the Individual Loss Policy (ILP). In all the examples the Optimal rule outperforms the other three showing the difficulty of creating a stopping rule that leads to losses as small as the optimal one.

In the first row of histograms on Figure 7.6 the results are related to the global loss function, and in the second one to the local loss. Note that the horizontal axis in each line is exactly the

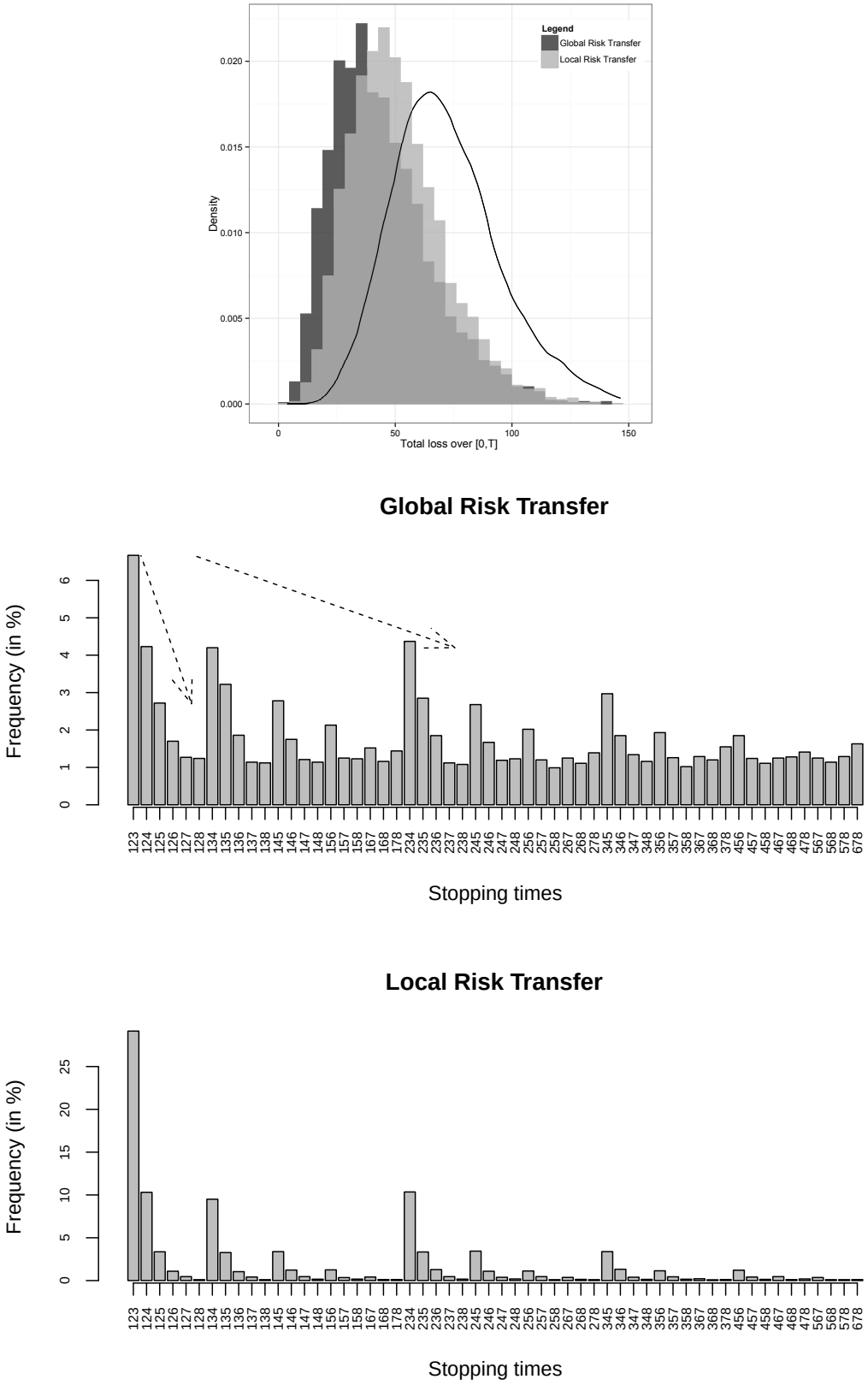

Figure 7.5: Comparison of the two objective functions using the Accumulated Loss Policy (ALP): (top) histograms of the total loss under the global objective function (dark grey), local objective function (light grey), no insurance case (solid line); (bottom) Multiple optimal stopping times under the two loss functions.

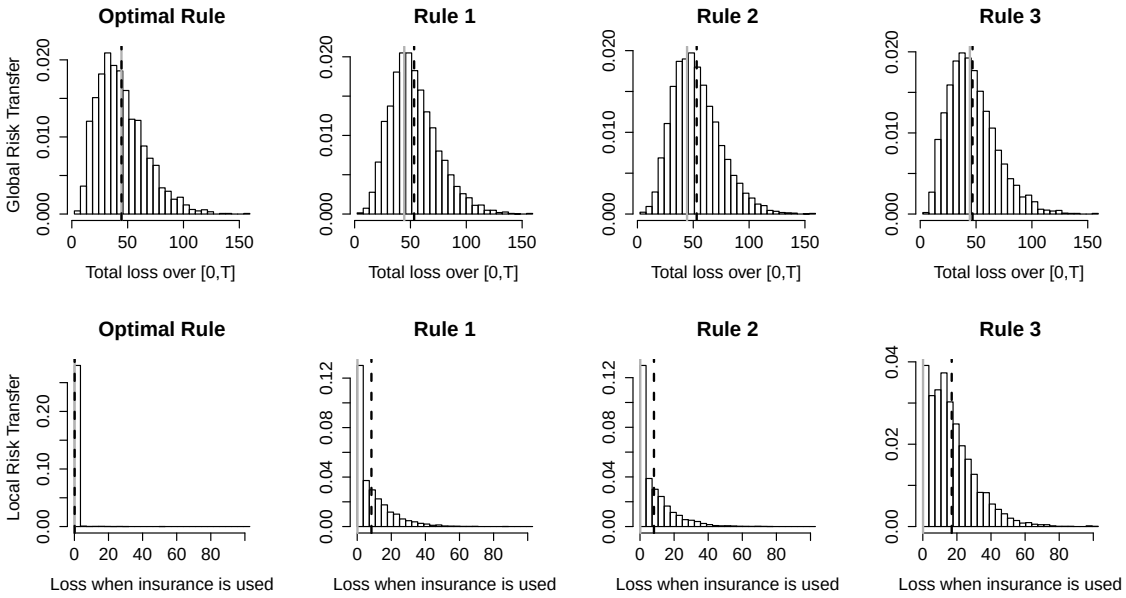

Figure 7.6: Histogram of losses under four different stopping rules for the ALP case with $(\lambda, \mu, \lambda_N) = (3, 2, 3)$ and $ALP = 10$.

objective function we are trying to minimize, precisely, $\sum_{\substack{t=1 \\ t \notin \{\tau_1, \ldots, \tau_k\}}}^{T} Z(t) + \sum_{j=1}^{k} \widetilde{Z}(\tau_j)$ for the global optimization and $\sum_{j=1}^{k} \widetilde{Z}(\tau_j)$ for the local one. In this figure the vertical dashed bar represents the average total loss under the different rules and the solid grey line is defined as

1. $\mathbb{E}[Z] \times T - v^{T,k}$, for the global optimization
2. $v^{T,k}$, for the local optimization.

These values must be understood as the expected loss under each of the two different gain functions and are easily derived from the definition of the gain functions and Theorem 7.3.5.

On Figure 7.7 we present the same comparison as in the second row of Figure 7.6 using the modelling proposed in Section 7.5.2.1, with parameters $(\lambda, \mu, \lambda_{\widetilde{N}}) = (3, 1, 4)$. For this simulation study the conclusion is similar to the one drawn from the ALP case, where the pre-defined stopping rules underperformed the multiple optimal rule.

7.7 Series expansion for the density of the insured process

Section 7.5 presented some combinations of Insurance Policies and LDA models that lead to closed form solutions for the multiple stopping rule. For the cases where analytical solutions can not be found, one alternative is to create a series expansion of the density of the insured process $\widetilde{Z}$ such that all the expectations necessary in Theorem 7.3.5 can be analytically calculated. In this Section we assume the first n moments of the distribution of the insured process $\widetilde{Z}$ are known and our objective is to minimize the local risk, but the calculations are also valid if we work with the global optimization problem (in this case one should use $Z - \widetilde{Z}$ instead of $\widetilde{Z}$).

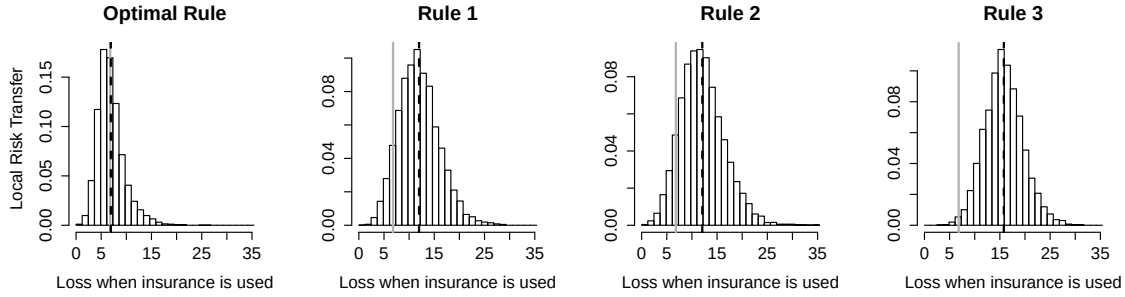

Figure 7.7: Histogram of losses under four different stopping rules for the ILP case with $(\lambda, \mu, \lambda_{\widetilde{N}}) = (3, 1, 4)$

Alternatively, one could proceed with some of the well established approximations of compound Poisson distributions. For a comparison of some approximation schemes, in the context of operational risk, the reader is referred to [Shevchenko, 2011, Chapter 3]. For a book-length discussion on these approximations, see [Willmot and Lin, 2001].

7.7.1 Gamma basis approximation

If the n-th first moments of the insured process $\widetilde{Z}$ can be calculated (either algebraically or numerically) and the support of the insured random variable is $[0, +\infty)$ one can use a series expansion of the density of $\widetilde{Z}$ in a Gamma basis. For notational convenience, define a new random variable $U = b\widetilde{Z}$, where $b = \frac{\mathbb{E}[\widetilde{Z}]}{Var[\widetilde{Z}]}$ and set $a = \frac{\mathbb{E}[\widetilde{Z}]^2}{Var[\widetilde{Z}]}$. Denoting by f_U the density of U the idea, as in the Gaussian case of a Gram-Charlier expansion, is to write f_U as

$$f_U(u) = g(u; a) \left[A_0 L_0^{(a)}(u) + A_1 L_1^{(a)}(u) + A_2 L_2^{(a)}(u) + \ldots \right]. \tag{7.10}$$

Since $supp(U) = supp(\widetilde{Z}) = [0, +\infty)$ we assume the kernel $g(\cdot\,; a)$ also has positive support (differently from the Gram-Charlier expansion, where $g(\cdot)$ is chosen as a Gaussian kernel). If $g(u; a) = \frac{u^{a-1}e^{-u}}{\Gamma(a)}$ i.e., a Gamma kernel with $shape = a$ and $scale = 1$, then the orthonormal polynomial basis (with respect to this kernel) is given by the Laguerre polynomials (in contrast to Hermite polynomials in the Gaussian case) defined as

$$L_n^{(a)}(u) = (-1)^n u^{1-a} e^{-u} \frac{d^n}{du^n}(u^{n+a-1}e^{-u}). \tag{7.11}$$

$L_0^{(a)}(u) =$	1
$L_1^{(a)}(u) =$	$u - a$
$L_2^{(a)}(u) =$	$u^2 - 2(a+1)u + (a+1)a$
$L_3^{(a)}(u) =$	$u^3 - 3(a+2)u^2 + 3(a+2)(a+1)u - (a+2)(a+1)a$
$L_4^{(a)}(u) =$	$u^4 - 4(a+3)u^3 + 6(a+3)(a+2)u^2 - 4(a+3)(a+2)(a+1)u + (a+3)(a+2)(a+1)a$

Table 7.2: The first five Laguerre polynomials

Remark 7.7.1. *Note that the definition of the Laguerre polynomials on Equation (7.11) is slightly different from the usual one, based on Rodrigues' formula*

$$\widetilde{L}_n^{(a)} = \frac{u^{-a}e^x}{n!}\frac{d^n}{du^n}\Big(e^{-x}x^{n+a}\Big),$$

but it is easy to check that

$$L_n^{(a)}(u) = n!(-1)^n\widetilde{L}_n^{(a-1)}.$$

From the orthogonality condition (see, for example, [Jackson, 1941] p. 184),

$$\int_0^{+\infty}\frac{x^{a-1}e^{-x}}{\Gamma(a)}L_n^{(a)}(x)L_m^{(a)}(x)dx = \begin{cases}\frac{n!\Gamma(a+n)}{\Gamma(a)},\ n=m,\\ 0,\ n\neq m\end{cases}$$

and using the fact that f_U can be written in the form of Equation (7.10) we find that

$$A_n = \frac{\Gamma(a)}{n!\Gamma(a+n)}\int_0^{+\infty} f_U(x)L_n^{(a)}(x)dx. \tag{7.12}$$

Then, using the characterization of A_n in (7.12) and the fact that $\mathbb{E}[U] = Var[U] = a$ we see that

$$\begin{aligned}
A_0 &= \int_0^{+\infty} f_U(x)L_0^{(a)}(x)dx = \int_0^{+\infty} f_U(x)dx = 1,\\
A_1 &= \int_1^{+\infty} f_U(x)L_1^{(a)}(x)dx = \int_0^{+\infty} f_U(x)(z-a)dx = 0,\\
A_2 &= \int_1^{+\infty} f_U(x)L_2^{(a)}(x)dx = \int_0^{+\infty} f_U(x)(z^2-2(a+1)z+(a+1)a)dx = 0.
\end{aligned}$$

Similar but lengthier calculations show that for $\mu_n = \mathbb{E}\left[(U-\mathbb{E}[U])^n\right]$, $n = 3, 4$,

$$A_3 = \frac{\Gamma(a)}{3!\Gamma(a+3)}(\mu_3 - 2a), \tag{7.13}$$

$$A_4 = \frac{\Gamma(a)}{4!\Gamma(a+4)}(\mu_4 - 12\mu_3 - 3a^2 + 18a). \tag{7.14}$$

Therefore, matching the first four moments, the density of the original random variable $\widetilde{Z}$ can be approximated as

$$f_{\widetilde{Z}}(z) = bf_U(u) \approx b\frac{u^{a-1}e^{-u}}{\Gamma(a)}\left[1 + A_3L_3^{(a)}(u) + A_4L_4^{(a)}(u)\right],$$

where $u = bz$, A_3 and A_4 are given, respectively, by (7.13) and (7.14) and the Laguerre polynomials can be found in Table 7.2. For additional details on the Gamma expansion we refer the reader to [Bowers Jr, 1966].

Since this expansion does not ensure positivity of the density at all points (it can be negative for particular choices of skewness and kurtosis) we adopt the approach discussed in [Jondeau and Rockinger, 2001] for the Gauss-Hermite Gramm Charlier case, modifying it to the Gamma-Laguerre setting. To find the region on the (μ_3, μ_4)-plane where $f_U(u)$ is positive for all u we first find the region where $f_U(u) = 0$, i.e.,

$$\frac{u^{a-1}e^{-u}}{\Gamma(a)}\big(1 + A_3L_3^{(a)}(u) + A_4L_4^{(a)}(u)\big) = 0. \tag{7.15}$$

For a fixed value u, we now want to find the set (μ_3, μ_4) as a function of u such that (7.15) remains zero for small variations on u. This set is given by (μ_3, μ_4) such that

$$\frac{d}{du}\left[\frac{u^{a-1}e^{-u}}{\Gamma(a)}\big(1 + A_3 L_3^{(a)}(u) + A_4 L_4^{(a)}(u)\big)\right] = 0. \tag{7.16}$$

We then rewrite Equations (7.15) and (7.15) as the following system of algebraic equations

$$\begin{cases} \mu_3 B_1(u) + \mu_4 B_2(u) + B_3(u) = 0 \\ \mu_3 B_1'(u) + \mu_4 B_2'(u) + B_3'(u) = 0, \end{cases}$$

where

$$B_1(u) = \frac{u^{a-1}e^{-u}}{\Gamma(a)}\left(\frac{\Gamma(a)}{3!\Gamma(a+3)}L_3^{(a)}(u) - 12\frac{\Gamma(a)}{4!\Gamma(a+4)}L_4^{(a)}(u)\right);$$

$$B_2(u) = \frac{u^{a-1}e^{-u}}{\Gamma(a)}\frac{\Gamma(a)}{4!\Gamma(a+4)}L_4^{(a)}(u);$$

$$B_3(u) = \frac{u^{a-1}e^{-u}}{\Gamma(a)}\left(1 - 2a\frac{\Gamma(a)}{3!\Gamma(a+3)}L_3^{(a)}(u) + \left(-3a^2 + 18a\right)\frac{\Gamma(a)}{4!\Gamma(a+4)}L_4^{(a)}(u)\right);$$

$$B_1'(u) = \left((a-1)u^{-1} - 1\right)B_1(u) + \frac{u^{a-1}e^{-u}}{\Gamma(a)}\left(\frac{\Gamma(a)}{3!\Gamma(a+3)}\frac{dL_3^{(a)}}{du}(u) - 12\frac{\Gamma(a)}{4!\Gamma(a+4)}\frac{dL_4^{(a)}}{du}(u)\right);$$

$$B_2'(u) = \left((a-1)u^{-1} - 1\right)B_2(u) + \frac{u^{a-1}e^{-u}}{\Gamma(a)}\left(\frac{\Gamma(a)}{4!\Gamma(a+4)}\frac{dL_4^{(a)}}{du}(u)\right);$$

$$B_3'(u) = \left((a-1)u^{-1} - 1\right)B_3(u) + \frac{u^{a-1}e^{-u}}{\Gamma(a)}\left(-2a\frac{\Gamma(a)}{3!\Gamma(a+3)}\frac{dL_3^{(a)}}{du}(u) + \left(-3a^2 + 18a\right)\frac{\Gamma(a)}{4!\Gamma(a+4)}\frac{dL_4^{(a)}}{du}(u)\right);$$

$$\frac{dL_3^{(a)}}{du}(u) = 3u^2 - 6(a+2)u + 3(a+2)(a+1);$$

$$\frac{dL_4^{(a)}}{du}(u) = 4u^3 - 12(a+3)u^2 + 12(a+3)(a+2)u - 4(a+3)(a+2)(a+1).$$

Therefore, solving this system we find the curve where the approximation stays positive for all u to be given by

$$\begin{cases} \mu_4(u) = \left(\frac{B_1'B_3}{B_1} - B_3'\right)\left(B_2' - \frac{B_1'B_2}{B_1}\right)^{-1} \\ \mu_3(u) = -\frac{1}{B_1}\left(\mu_4(u)B_2 + B_3\right) \end{cases}, \text{ for } u \in [0, +\infty). \tag{7.17}$$

As an illustration, Figure 7.8 presents (on the left) the histogram of the loss process $Z = \sum_{n=1}^{N} X_n$ for $X \sim LN(\mu = 1, \sigma = 0.8)$ and $N \sim Poi(\lambda_N = 2)$ and in red the Gamma approximation using the first four moments of Z. On the right it is presented the graph of the region where the density is positive for all values of u, given by equation 7.17. The grey area was calculated numerically, for all combinations in a fine grid on the plane (μ_3, μ_4) it was tested if the density became negative in some point z. Grey points indicate the density is strictly positive. The dark dot indicates the third and fourth moments in the Log-Normal example and since it lies inside the positivity area we can ensure this approximation is strictly positive for all values of z.

If the the third and fourth moments of the chosen model lied outside the permitted area one could chose $\widehat{\mu}_3$ and $\widehat{\mu}_4$ as the estimates that minimize some constrained optimization problem, for instance, the Maximum Likelihood Estimator (using $f_U(u; \mu_3, \mu_4) =$

$\frac{u^{a-1}e^{-u}}{\Gamma(a)}\left[1+A_3L_3^{(a)}(u)+A_4L_4^{(a)}(u)\right]$ as the likelihood). The constrained region is clearly given by a segment of the curve in equation 7.17 and the endpoints can be found using a root-search method checking for which values of u the curve in Figure 7.8 touches the grey area.

Given the approximation of f_U, and consequently of $f_{\widetilde{Z}}$, one can easily calculate the optimal multiple stopping rule, since $\mathbb{E}[\widetilde{Z}]$ is assumed to be known and $\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}]$ can be calculated as follows.

Lemma 7.7.2. *If $G\sim Gamma(a,1)$, ie, $f_G(x)=\frac{x^{a-1}e^{-x}}{\Gamma(a)}$ then, similarly to Lemma 7.4.4 the following property holds*

$$xf_G(x;\ a,1)\equiv af_G(x;\ a+1,1). \tag{7.18}$$

Using this notation we can rewrite the approximation of $\widetilde{Z}$ as

$$f_{\widetilde{Z}}(z)\approx f_G(bz;a,1)A_1^*+f_G(bz;a+1,1)A_2^*+f_G(bz;a+2,1)A_3^*+f_G(bz;a+3,1)A_4^*+f_G(bz;a+4,1)A_5^*,$$

where $A_1^*=\left(1-\frac{\Gamma(a+3)}{\Gamma(a)}A_3+\frac{\Gamma(a+4)}{\Gamma(a)}A_4\right)b,\ A_2^*=\left(3\frac{\Gamma(a+3)}{\Gamma(a)}A_3-4\frac{\Gamma(a+4)}{\Gamma(a)}A_4\right)b,$
$A_3^*=\left(-3\frac{\Gamma(a+3)}{\Gamma(a)}A_3+6\frac{\Gamma(a+4)}{\Gamma(a)}A_4\right)b,\ \ A_4^*=\left(\frac{\Gamma(a+3)}{\Gamma(a)}A_3-4\frac{\Gamma(a+4)}{\Gamma(a)}A_4\right)b, A_5^*=\left(\frac{\Gamma(a+4)}{\Gamma(a)}A_4\right)b.$

Then, we can calculate the other main ingredient of Theorem 7.3.5, namely

$$\begin{aligned}
\mathbb{E}[\min\{c_1+\widetilde{Z},\ c_2\}] &= \int_0^{+\infty}\min\{c_1+z,\ c_2\}f_{\widetilde{Z}}(z)dz\\
&= \int_0^{+\infty}\left((c_1+z)\mathbb{1}_{\{c_1+z<c_2\}}+c_2\mathbb{1}_{\{c_1+z\geq c_2\}}\right)f_{\widetilde{Z}}(z)dz\\
&= \int_0^{c_2-c_1}zf_{\widetilde{Z}}(z)dz+c_1\int_0^{c_2-c_1}f_{\widetilde{Z}}(z)dz+c_2\int_{c_2-c_1}^{+\infty}f_{\widetilde{Z}}(z)dz\\
&= a\sum_{k=1}^{5}F_G(b(c_2-c_1);a+k,1)A_k^*+c_1\sum_{k=1}^{5}F_G(b(c_2-c_1);a-1+k,1)A_k^*\\
&+c_2\sum_{k=1}^{5}\overline{F}_G(b(c_2-c_1);a-1+k,1)A_k^*.
\end{aligned}$$

7.8 Conclusions and final remarks

In this chapter we studied some properties of an insurance product where its owner has the right to choose which of the next k years the issuer should mitigate its annual losses. For two different forms of mitigation we presented closed form solutions for the exercising strategy that minimized (on average) the sum of all annual losses in the next T years. This model assumed a "moderate tail" for the severity of the losses the owner incurs, namely a Poisson-Inverse Gaussian LDA model.

Although it is assumed the company already holds the proposed contract, the company can use the analysis presented on Figure 7.5 as a proxy for the price of the insurance product. The value, from the company's point of view, of the insurance product should be the expected difference (under the natural probability) of the losses that would be incurred without the product and the losses incurred using the product in the most profitable way (for the buyer),

$$\mathbb{E}\left[\sum_{t=1}^{L}Z(t)-\left(\sum_{t\notin\{\tau_1,\ldots,\tau_k\}}^{L}Z(t)+\sum_{j=1}^{l}\widetilde{Z}(\tau_j)\right)\right].$$

It must also be said this price does not include the premium asked by the insurance company and also does not take into consideration the fact that external insurance companies will not have access to the models used by the company but it can still be a valuable proxy.

An alternative to the results presented in Section 7.7 can involve the use of a Monte Carlo method. If there exists a mechanism to sample from the severity distribution one can easily create a sample of the insured process $\widetilde{Z}$ and use this sample to calculate all the necessary expectations on Theorem 7.3.5. The advantage of this approach is that one can handle any combination of severity distribution and insurance policy, but it can be extremely time consuming and the variance of the estimative can be prohibitive. It is important to note the sampling of the severity can be made *offline*, i.e., the same sample should be used to calculate all the integrals. Another alternative to solve the optimal multiple stopping problem is the usage of an extended version of the so-called Least-Square Monte Carlo method, first presented in [Longstaff and Schwartz, 2001] and extended to the multiple stopping scenario in [Bender and Schoenmakers, 2006] (see also [Bender et al., 2013] for recent related work).

Regarding the results presented in Theorem 7.5.5 and 7.5.2 the truncation point for the infinite sums can be chosen to be much larger than the expected number of losses (parameter λ_N), since the summands are composed by a p.m.f. of a Poisson r.v. (which presents exponential decay) and a bounded term (difference of c.d.f.'s times constants).

From a critical perspective, it should be understood the policies introduced in this chapter are to be seen as a first attempt to reduce the price of operational insurance and still lack some real features. For example, neither the Individual Loss Policy (ILP) nor the Accumulated Loss Policy include deductibles. One possible extension would be to study the following adjusted structure for ILP: $\widetilde{Z} = Z - \min\{(Z - d)_+, l)$. The adjusted ILP structure would tackle the possible issue of moral hazard on the original ILP, where small claims (smaller than k) are always paid in full, without any sort of deductible. Another important factor not taken into consideration is the impact of the insurance policies in the distribution of the liabilities. For the ILP and ALP we do not discuss the impact of these policies, for example, in the capital requirements. Along these lines, the decision criteria, i.e., minimization of the expected loss, does not take into account the shape of the resulting distribution either. The optimization problem with respect to the whole distribution (in order to see the impact in capital requirements, for example) would be a much harder than just analyzing the expected values, and is left as a future direction of research.

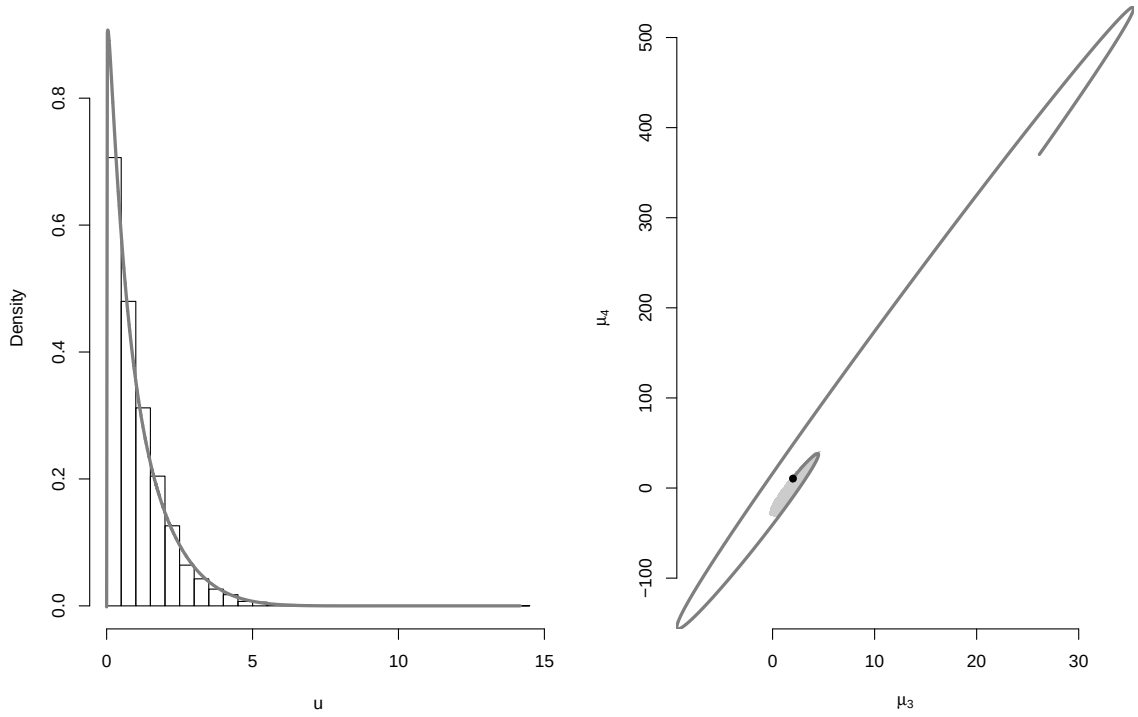

Figure 7.8: (Left) Histogram of the loss process $Z = \sum_{n=1}^{N} X_n$ for $X \sim LN(\mu = 1, \sigma = 0.8)$ and $N \sim Poi(\lambda_N = 2)$ and in red the Gamma approximation using the first four moments of Z. (Right) The graph of the region where the density is positive for all values of z.

Chapter 8

Conclusions and future work

The recent regulatory changes in the financial/actuarial landscape are leading to stricter capital requirements, making it even more important to develop tools to understand and manage the company's risks. This thesis provides quantitative tools to be used for this purpose. In particular, within the contexts of operational and actuarial risk management we discuss the problem of allocating the company's capital and also provide a new class of insurance products to help mitigating operational risk losses.

Regarding the capital allocation problem we introduce novel simulation algorithms based on Sequential Monte Carlo (SMC) methods, which can be applied to a wide range of models. In a scenario where no parameter uncertainty is present, we are able to perform the sampling procedure in a bounded space, which leads to good performance gains. In order to optimize the algorithm even further, one possible direction is to make use of the geometric properties of the constrained region, such as its curvature, as discussed in Section 4.1.1.

Parameter uncertainty is dealt with in the case of allocation of actuarial risks in Chapters 5 and 6, where we introduce a marginalized and a conditional model. For allocations under the conditional model we use a pseudo-marginal SMC algorithm, which involves the estimation, via simple Monte Carlo, of an intractable density. In light of the recent results from [Deligiannidis et al., 2015] and [Dahlin et al., 2015] it worth analyzing the impact of introducing correlated random variables when estimating the density $f_{\overline{Z}}$.

On a different direction, another alternative to be explored is to calculate a deterministic approximation to the unknown density $f_{\overline{Z}}$. When used within the SMC procedure this approximation induces some bias in the allocations, but may substantially reduce the computational cost, as the approximation is computed only once (outside the SMC loop). This strategy has been tested, for example, in [Everitt et al., 2016] and [McGree et al., 2015].

In order to make comparisons between the marginalized and conditional approaches simpler, we perform the sampling procedure of the later in the original space, instead of using the inverse c.d.f. transformation from Chapter 4. Although the copula implicitly defined in $f_{\overline{\boldsymbol{Z}}}$ cannot be unbiasedly estimated, a deterministic approximation can be derived, which would allow us to use the strategy discussed above. The result based on the deterministic approximation of the

(implicit) copula of $f_{\overline{\boldsymbol{Z}}}$ could then be compared with the SMC procedure in the (explicitly defined) copula space for f_Z.

Another possible extension is related to the risk measure used in the capital calculation. From the recent derivation of *expectiles* allocations in [Emmer et al., 2015] it is clear that similar SMC methods to those proposed in this work can be used, as similar conditional expectations are involved.

The SMC methodology presented in Chapters 4 and 5 could also accommodate the sensitivity calculations performed in [Tsanakas and Millossovich, 2016]. In this work, sensitivity analysis is based on quantities such as $\mathrm{ES}_\alpha(g(\mathbf{X}))$, where g is a non-linear function and $\mathbf{X}$ is a high dimensional vector (with several hundreds of variables). The efficiency of the SMC when compared to naive MC will depend on two important factors. First, as seen in the Gumbel copula example in Figure 4.8, the SMC method suffers from the so called "curse of dimensionality", where the relative bias increase and the variance reduction decreases as the dimension of $\mathbf{X}$ increases. The second factor is the cost of computing the function g. As discussed in Chapter 4, the theoretical computational gain of the SMC method over the naive MC is of the order $(1-\alpha)/T$. Since the naive MC method doesn't need to compute g, the SMC can be expected to present some efficiency gains whenever the cost of computing g for each sample is lower than than $(1-\alpha)/T$.

Related to the insurance product proposed in Chapter 7, the recent frustrating sales of "operational risk bonds" by Credit Suisse (see [Foerster and Beardsworth, 2016]), when the issuers were said to sell as little as 1/3 of the initially planned amount, show that the lack of flexible operational risk transfer strategies may be preventing companies from getting the coverage they want/need. In this regard, the class of insurance products introduced in Chapter 7 may provide a viable and affordable alternative to products that keep the company insured for the whole policy's life-span.

Bibliography

[Aas et al., 2009] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

[Acerbi, 2002] Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

[Acerbi and Székely, 2014] Acerbi, C. and Székely, B. (2014). Back-testing expected shortfall. *Risk*, page 76.

[Acerbi and Tasche, 2002] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

[Allen et al., 2009] Allen, L., Boudoukh, J., and Saunders, A. (2009). *Understanding market, credit, and operational risk: the value at risk approach*. Wiley.

[Andrieu et al., 2010] Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.

[Andrieu and Roberts, 2009] Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(1):697–725.

[Andrieu and Vihola, 2015] Andrieu, C. and Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability*, 25(2):1030–1077.

[Anonymous, 2009] Anonymous (2009). In defence of the Gaussian copula. Online: `http://www.economist.com/blogs/freeexchange/2009/04/in_defense_of_copula` Accessed: 23/Aug/2016.

[Arbenz et al., 2014] Arbenz, P., Cambou, M., and Hofert, M. (2014). An importance sampling algorithm for copula models in insurance. *arXiv preprint arXiv:1403.4291*.

[Artzner et al., 1999] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228.

[Asimit et al., 2013] Asimit, A. V., Vernic, R., and Zitikis, R. (2013). Evaluating risk measures and capital allocations based on multi-losses driven by a heavy-tailed background risk: The multivariate Pareto-II model. *Risks*, 1(1):14–33.

[Au and Beck, 2001] Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277.

[Aubin, 2007] Aubin, J.-P. (2007). *Mathematical methods of game and economic theory*. Courier Corporation.

[Bargès et al., 2009] Bargès, M., Cossette, H., and Marceau, E. (2009). TVaR-based capital allocation with copulas. *Insurance: Mathematics and Economics*, 45(3):348–361.

[Bazzarello et al., 2006] Bazzarello, D., Crielaard, B., Piacenza, F., and Soprano, A. (2006). Modeling insurance mitigation on operational risk capital. *Journal of Operational risk*, 1(1):57–65.

[BCBS, 1988] BCBS (1988). International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements.

[BCBS, 2003a] BCBS (2003a). A brief history of the Basel committee. Technical report, Bank for International Settlements.

[BCBS, 2003b] BCBS (2003b). A glossary of terms used in payments and settlement systems. Technical report, Bank for International Settlements.

[BCBS, 2006] BCBS (2006). International convergence of capital measurement and capital standards – A revised framework (comprehensive version). Technical report, Bank for International Settlements.

[BCBS, 2010a] BCBS (2010a). Basel III: a global regulatory framework for more resilient banks and baning systems. Technical report, Bank for International Settlements.

[BCBS, 2010b] BCBS (2010b). Basel III: International framework for liquidity risk measurement, standards and monitoring. Technical report, Bank for International Settlements.

[BCBS, 2014] BCBS (2014). Operational risk - revisions to the simpler approaches. Technical report, Bank for International Settlements.

[BCBS, 2016a] BCBS (2016a). Minimum capital requirements for market risk. Technical report, Bank for International Settlements.

[BCBS, 2016b] BCBS (2016b). Standardised measurement approach for operational risk. Technical report, Bank for International Settlements.

[Beaumont, 2003] Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160.

[Bender and Schoenmakers, 2006] Bender, C. and Schoenmakers, J. (2006). An iterative method for multiple stopping: convergence and stability. *Advances in applied probability*, pages 729–749.

[Bender et al., 2013] Bender, C., Schoenmakers, J., and Zhang, J. (2013). Dual representations for general multiple stopping problems. *Mathematical Finance*.

[Beskos et al., 2006] Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

[Bignozzi, 2012] Bignozzi, V. (2012). *Contributions to solvency risk measurement*. PhD thesis, City University.

[Birge et al., 2012] Birge, J. R., Chang, C., and Polson, N. G. (2012). Split sampling: Expectations, normalisation and rare events. *arXiv preprint arXiv:1212.0534*.

[Boonen et al., 2016] Boonen, T. J., Tsanakas, A., and Wuthrich, M. V. (2016). Capital allocation for insurance portfolios with non-linear risk aggregation. *Available at SSRN*.

[Bowers Jr, 1966] Bowers Jr, N. L. (1966). Expansion of probability density functions as a sum of gamma densities with applications in risk theory. *Trans. Soc. Actuaries*, 18:125–137.

[Brandts, 2004] Brandts, S. (2004). Operational risk and insurance: quantitative and qualitative aspects. *Paper in progress, Goethe University, Frankfurt*.

[Buch and Dorfleitner, 2008] Buch, A. and Dorfleitner, G. (2008). Coherent risk measures, coherent capital allocations and the gradient allocation principle. *Insurance: Mathematics and Economics*, 42(1):235–242.

[Cai et al., 2015] Cai, J.-J., Einmahl, J. H., Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):417–442.

[Capéraà and Genest, 1993] Capéraà, P. and Genest, C. (1993). Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *Journal of Nonparametric Statistics*, 2(2):183–194.

[Carmona and Crépey, 2010] Carmona, R. and Crépey, S. (2010). Particle methods for the estimation of credit portfolio loss distributions. *International Journal of Theoretical and Applied Finance*, 13(04):577–602.

[Carmona et al., 2009] Carmona, R., Fouque, J.-P., and Vestal, D. (2009). Interacting particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, 13(4):613–633.

[Carmona and Touzi, 2008] Carmona, R. and Touzi, N. (2008). Optimal multiple stopping and valuation of swing options. *Mathematical Finance*, 18(2):239–268.

[Cérou et al., 2012] Cérou, F., Del Moral, P., Furon, T., and Guyader, A. (2012). Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808.

[Cérou et al., 2006] Cérou, F., Del Moral, P., Le Gland, F., and Lezaud, P. (2006). Genetic genealogical models in rare event analysis. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, 1:181–203.

[Cérou and Guyader, 2007] Cérou, F. and Guyader, A. (2007). Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443.

[Cérou et al., 2005] Cérou, F., LeGland, F., Moral, P. D., and Lezaud, P. (2005). Limit theorems for the multilevel splitting algorithm in the simulation of rare events. In *Simulation Conference, 2005 Proceedings of the Winter*, pages 10–pp. IEEE.

[Chernobai et al., 2008] Chernobai, A. S., Rachev, S. T., and Fabozzi, F. J. (2008). *Operational risk: a guide to Basel II capital requirements, models, and analysis*, volume 180. John Wiley & Sons.

[Cherubini et al., 2004] Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.

[Chopin, 2004] Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *The Annals of Statistics*, 32(6):2385–2411.

[Chopin et al., 2013] Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). SMC^2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):397–426.

[Cont et al., 2010] Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10(6):593–606.

[Cossette et al., 2013] Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013). Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3):560–572.

[Creal, 2012] Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, 31(3):245–296.

[Crisan and Doucet, 2002] Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *Signal Processing, IEEE Transactions on*, 50(3):736–746.

[Cruz et al., 2015] Cruz, M. G., Peters, G. W., and Shevchenko, P. V. (2015). *Fundamental Aspects of Operational Risk Modeling and Insurance Analytics: A Handbook of Operational Risk*. John Wiley & Sons.

[Dahlin et al., 2015] Dahlin, J., Lindsten, F., Kronander, J., and Schön, T. B. (2015). Accelerating pseudo-marginal metropolis-hastings by correlating auxiliary variables. *arXiv preprint arXiv:1511.05483*.

[Dall'Aglio et al., 1991] Dall'Aglio, G., Kotz, S., and Salinetti, G. (1991). *Advances in probability distributions with given marginals: beyond the copulas*. Kluwer Academic Publishers, London.

[Daníelsson et al., 2013] Daníelsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., and de Vries, C. G. (2013). Fat tails, var and subadditivity. *Journal of econometrics*, 172(2):283–291.

[Das and Scism, 2016] Das, A. and Scism, L. (2016). New from Credit Suisse: Bonds for self-inflicted catastrophes. Online: `http://www.wsj.com/articles/new-fromcredit-suisse-bonds-for-self-inflicted-catastrophes-1463419276`, Accessed: 14/Jul/2016.

[Das et al., 2013] Das, B., Embrechts, P., and Fasen, V. (2013). Four theorems and a financial crisis. *International Journal of Approximate Reasoning*, 54(6):701–716.

[De Luca and Rivieccio, 2012] De Luca, G. and Rivieccio, G. (2012). Multivariate tail dependence coefficients for Archimedean copulae. In *Advanced Statistical Methods for the Analysis of Large Data-Sets*, pages 287–296. Springer.

[Del Moral, 2004] Del Moral, P. (2004). *Feynman-Kac Formulae*. Springer.

[Del Moral et al., 2006] Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

[Del Moral et al., 2013] Del Moral, P., Peters, G. W., and Vergé, C. (2013). An introduction to stochastic particle integration methods: with applications to risk and insurance. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 39–81. Springer.

[Delbaen, 2000] Delbaen, F. (2000). Coherent risk measures. PISA Lecture notes.

[Deligiannidis et al., 2015] Deligiannidis, G., Doucet, A., Pitt, M. K., and Kohn, R. (2015). The correlated pseudo-marginal method. *arXiv preprint arXiv:1511.04992*.

[Denault, 2001] Denault, M. (2001). Coherent allocation of risk capital. *Journal of risk*, 4(1):1–34.

[Denneberg, 1994] Denneberg, D. (1994). *Non-additive measure and integral*, volume 27. Springer Science & Business Media.

[Denuit and Dhaene, 2003] Denuit, M. and Dhaene, J. (2003). Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. *Belgian Actuarial Bulletin*, 3(1):22–27.

[Devroye and Letac, 2015] Devroye, L. and Letac, G. (2015). Copulas with prescribed correlation matrix. In *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, pages 585–601. Springer.

[Dhaene et al., 2008] Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A., and Vanduffel, S. (2008). Some results on the CTE-based capital allocation rule. *Insurance: Mathematics and Economics*, 42(2):855–863.

[Dhaene et al., 2012] Dhaene, J., Tsanakas, A., Valdez, E. A., and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28.

[Donnelly and Embrechts, 2010] Donnelly, C. and Embrechts, P. (2010). The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. *Astin Bulletin*, 40(01):1–33.

[Doucet et al., 2001] Doucet, A., De Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer.

[Doucet and Johansen, 2009] Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12:656–704.

[Doucet et al., 2015] Doucet, A., Pitt, M., Deligiannidis, G., and Kohn, R. (2015). Efficient implementation of markov chain monte carlo when using an unbiased likelihood estimator. *Biometrika*, page asu075.

[Duane et al., 1987] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, 195(2):216–222.

[Durante et al., 2015] Durante, F., Puccetti, G., and Scherer, M. (2015). Building bridges between mathematics, insurance and finance. *Dependence Modeling*, 3(1).

[EIOPS, 2010] EIOPS (2010). QIS 5 techical specifications. Technical report, European Insurance and Occupational Pensions Authority.

[Embrechts, 2008] Embrechts, P. (2008). Statistics* and quantitative risk management (* including computation probability). Presentation: `https://people.math.ethz.ch/~embrecht/ftp/Ottawa2008_Dec_english.pdf` Accessed: 23/Aug/2016.

[Embrechts et al., 2003] Embrechts, P., Lindskog, F., and McNeil, A. (2003). *Handbook of Heavy Tailed Distributions in Finance*, chapter Modelling dependence with copulas, pages 329–384. Elsevier, North–Holland.

[Embrechts et al., 2002] Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223.

[Embrechts et al., 2015] Embrechts, P., Wang, R., et al. (2015). Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling*, 3(1):126–140.

[Emmer et al., 2015] Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, 18(2):31–60.

[Escher, 1932] Escher, F. (1932). On the probability function in the collective theory of risk. *Skand. Aktuarie Tidskr.*, 15:175–195.

[Everitt et al., 2016] Everitt, R. G., Johansen, A. M., Rowing, E., and Evdemon-Hogan, M. (2016). Bayesian model comparison with un-normalised likelihoods. *Statistics and Computing*, pages 1–20.

[Finke, 2015] Finke, A. (2015). *On Extended State-Space Constructions for Monte Carlo Methods*. PhD thesis, University of Warwick.

[FINMA, 2007] FINMA (2007). Technical document on the Swiss solvency test. Technical report.

[FINMA, 2016] FINMA (2016). Standardmodell Schadenversicherung. Online: File `sst-template-nonlife_2016.xlsx`, tab `NL_Insurance Risk`, address `https://www.finma.ch/de/~/media/finma/dokumente/dokumentencenter/myfinma/2ueberwachung/sst/standard-model-nonlife-2016.zip?la=de`. Accessed: 13/Jul/2016.

[Foerster and Beardsworth, 2016] Foerster, J.-H. and Beardsworth, T. (2016). Credit Suisse's 'rogue trader' bond falls short on risk. Online: `http://www.bloomberg.com/news/articles/2016-05-26/credit-suisse-s-rogue-trader-bond-falls-short-as-risk-spurned`, Accessed: 14/Jul/2016.

[Folks and Chhikara, 1978] Folks, J. and Chhikara, R. (1978). The inverse Gaussian distribution and its statistical application – a review. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):263–289.

[Föllmer and Schied, 2002] Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447.

[Franzetti, 2011] Franzetti, C. (2011). *Operational risk modelling and management*. Taylor & Francis US.

[Fredricks and Nelsen, 2007] Fredricks, G. A. and Nelsen, R. B. (2007). On the relationship between Spearman's *rho* and Kendall's *tau* for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, 137(7):2143–2150.

[Frees and Valdez, 1998] Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American actuarial journal*, 2(1):1–25.

[Frittelli and Gianin, 2002] Frittelli, M. and Gianin, E. R. (2002). Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486.

[Fulop and Li, 2013] Fulop, A. and Li, J. (2013). Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics*, 176(2):146–161.

[Furman and Landsman, 2005] Furman, E. and Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, 37(3):635–649.

[Gamerman and Lopes, 2006] Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press.

[Garvels, 2000] Garvels, M. (2000). *The splitting in rare event estimation*. PhD thesis, Ph. D. thesis, University of Twente, Twente, Netherlands.

[Gelman and Meng, 1998] Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741.

[Geweke, 1989] Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339.

[Gilks and Berzuini, 2001] Gilks, W. R. and Berzuini, C. (2001). Following a moving target Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):127–146.

[Gisler, 2006] Gisler, A. (2006). The estimation error in the chain-ladder reserving method: a bayesian approach. *Astin Bulletin*, 36(02):554–565.

[Gisler and Wüthrich, 2008] Gisler, A. and Wüthrich, M. V. (2008). Credibility for the chain ladder reserving method. *Astin Bulletin*, 38(02):565–600.

[Glasserman, 2005] Glasserman, P. (2005). Measuring marginal risk contributions in credit portfolios. *Journal of Computational Finance*, 9(2):1.

[Glasserman et al., 1996] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. (1996). Splitting for rare event simulation: analysis of simple cases. In *Proceedings of the 28th conference on Winter simulation*, pages 302–308. IEEE Computer Society.

[Glasserman et al., 1999] Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. (1999). Multilevel splitting for estimating rare event probabilities. *Operations Research*, 47(4):585–600.

[Gneiting, 2011] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

[Gouriéroux et al., 2000] Gouriéroux, C., Laurent, J. P., and Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3):225–245.

[Hastings, 1970] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

[Hofert, 2010] Hofert, J. M. (2010). *Sampling Nested Archimedean Copulas: With Applications to CDO Pricing*. Suedwestdeutscher Verlag fuer Hochschulschriften.

[Hofert et al., 2014] Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2014). *copula: Multivariate Dependence with Copulas*. R package version 0.999-9.

[Hull, 2012] Hull, J. (2012). *Risk Management and Financial Institutions + Web Site*, volume 733. John Wiley & Sons.

[Jackson, 1941] Jackson, D. (1941). *Fourier series and orthogonal polynomials*. Courier Dover Publications.

[Jaillet et al., 2004] Jaillet, P., Ronn, E. I., and Tompaidis, S. (2004). Valuation of commodity-based swing options. *Management science*, 50(7):909–921.

[Jasra and Del Moral, 2011] Jasra, A. and Del Moral, P. (2011). Sequential monte carlo methods for option pricing. *Stochastic analysis and applications*, 29(2):292–316.

[Joe, 1997] Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

[Joe, 2014] Joe, H. (2014). *Dependence modeling with copulas*. CRC Press.

[Johansen and Evers, 2007] Johansen, A. M. and Evers, L. (2007). *Monte Carlo Methods*. (unpublished).

[Johansen et al., 2006] Johansen, A. M., Moral, P. D., and Doucet, A. (2006). Sequential monte carlo samplers for rare events. In *Proceedings of the 6th International Workshop on Rare Event Simulation*, pages 256–267.

[Jondeau and Rockinger, 2001] Jondeau, E. and Rockinger, M. (2001). Gram–Charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483.

[Jørgensen, 1982] Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*, volume 9 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

[Kahn and Harris, 1951] Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards applied mathematics series*, 12:27–30.

[Kalkbrener, 2005] Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3):425–437.

[Kennedy and Kuti, 1985] Kennedy, A. and Kuti, J. (1985). Noise without noise: a new monte carlo method. *Physical review letters*, 54(23):2473.

[Kroese et al., 2013] Kroese, D. P., Rubinstein, R. Y., and Glynn, P. W. (2013). The cross-entropy method for estimation. *Handbook of Statistics: Machine Learning: Theory and Applications*, 31:19–34.

[Krupskii and Joe, 2013] Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

[Kruskal, 1958] Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53(284):814–861.

[Künsch, 2005] Künsch, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Annals of Statistics*, 33(5):1983–2021.

[Landsman and Valdez, 2003] Landsman, Z. M. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

[L'Ecuyer et al., 2006] L'Ecuyer, P., Demers, V., and Tuffin, B. (2006). Splitting for rare-event simulation. In *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*, pages 137–148. IEEE.

[Li, 2000] Li, D. X. (2000). On default correlation: A copula function approach. *Journal of Fixed Income, March*.

[Lin et al., 2000] Lin, L., Liu, K., and Sloan, J. (2000). A noisy monte carlo algorithm. *Physical Review D*, 61(7):074505.

[Liu and Chen, 1998] Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044.

[Longstaff and Schwartz, 2001] Longstaff, F. A. and Schwartz, E. S. (2001). Valuing american options by simulation: A simple least-squares approach. *Review of Financial studies*, 14(1):113–147.

[Mack, 1993] Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, 23(02):213–225.

[McGree et al., 2015] McGree, J. M., Drovandi, C. C., White, G., and Pettitt, A. N. (2015). A pseudo-marginal sequential Monte Carlo algorithm for random effects models in Bayesian sequential design. *Statistics and Computing*, 26(5):1–16.

[McLeish, 2010] McLeish, D. L. (2010). Bounded relative error importance sampling and rare event simulation. *Astin Bulletin*, 40(1):377–398.

[McNeil, 2008] McNeil, A. J. (2008). Sampling nested archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

[McNeil et al., 2010] McNeil, A. J., Frey, R., and Embrechts, P. (2010). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.

[McNeil et al., 2015] McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.

[Merz and Wüthrich, 2015] Merz, M. and Wüthrich, M. V. (2015). Claims run-off uncertainty: the full picture. *Available at SSRN 2524352, version of 3/Jul/2015*.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

[Mikosch, 2006] Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1):3–20.

[Mildenhall, 2004] Mildenhall, S. J. (2004). A note on the Myers and Read capital allocation formula. *North American Actuarial Journal*, 8(2):32–44.

[Mildenhall, 2006] Mildenhall, S. J. (2006). Actuarial geometry. In *Risk Theory Seminar*.

[Murray, 2013] Murray, L. M. (2013). Bayesian state-space modelling on high-performance hardware using libbi. *arXiv preprint arXiv:1306.3277*.

[Neal, 2003] Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–741.

[Neal, 2011] Neal, R. M. (2011). *Handbook of Markov Chain Monte Carlo*, volume 2, chapter MCMC using Hamiltonian dynamics, pages 113–162. CRC Press.

[Nelsen, 2007] Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

[Nikolaev and Sofronov, 2007] Nikolaev, M. and Sofronov, G. (2007). A multiple optimal stopping rule for sums of independent random variables. *Discrete Mathematics and Applications dma*, 17(5):463–473.

[Oh and Patton, 2013] Oh, D. H. and Patton, A. J. (2013). Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502):689–700.

[Okhrin and Ristig, 2014] Okhrin, O. and Ristig, A. (2014). Hierarchical archimedean copulae: The hac package. *Journal of Statistical Software*, 58(4):1–20.

[Panjer, 2001] Panjer, H. H. (2001). *Measureement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates*. University of Waterloo, Institute of Insurance and Pension Research.

[Peters, 2005] Peters, G. W. (2005). Topics in sequential Monte Carlo samplers. *M. Sc., University of Cambridge, Department of Engineering*.

[Peters et al., 2011] Peters, G. W., Byrnes, A. D., and Shevchenko, P. V. (2011). Impact of insurance for operational risk: Is it worthwhile to insure or be insured for severe losses? *Insurance: Mathematics and Economics*, 48(2):287–303.

[Peters et al., 2016a] Peters, G. W., Shevchenko, P. V., Hassani, B., and Chapelle, A. (2016a). Standardized measurement approach for operational risk: Pros and cons. *Available at SSRN 2789006*.

[Peters et al., 2013] Peters, G. W., Targino, R. S., and Shevchenko, P. V. (2013). Understanding operational risk capital approximations: First and second orders. *Journal of Governance and Regulation*, 2(3):58–78.

[Peters et al., 2016b] Peters, G. W., Targino, R. S., and Wuthrich, M. V. (2016b). Bayesian modelling and allocation of insurance risks. Manuscript in preparation.

[Peters et al., 2017] Peters, G. W., Targino, R. S., and Wüthrich, M. V. (2017). Full bayesian analysis of claims reserving uncertainty. *Insurance: Mathematics and Economics*, in press.

[Pitt et al., 2012] Pitt, M. K., dos Santos Silva, R., Giordani, P., and Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151.

[Rambharat and Brockwell, 2010] Rambharat, B. R. and Brockwell, A. E. (2010). Sequential monte carlo pricing of american-style options under stochastic volatility models. *The Annals of Applied Statistics*, pages 222–265.

[Reinhart and Rogoff, 2009] Reinhart, C. M. and Rogoff, K. (2009). *This time is different: eight centuries of financial folly*. princeton university press.

[Ristic et al., 2004] Ristic, B., Arulampalam, S., and Gordon, N. (2004). *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house.

[Robert and Casella, 2004] Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Science & Business Media.

[Salmon, 2009] Salmon, F. (2009). Recipe for disaster: the formula that killed Wall Street. *Wired Magazine*, 17(3).

[Sen et al., 2016] Sen, D., Jasra, A., and Zhou, Y. (2016). Some contributions to sequential monte carlo methods for option pricing. *arXiv preprint arXiv:1608.03352*.

[Sherlock et al., 2015] Sherlock, C., Thiery, A. H., Roberts, G. O., Rosenthal, J. S., et al. (2015). On the efficiency of pseudo-marginal random walk metropolis algorithms. *The Annals of Statistics*, 43(1):238–275.

[Shevchenko, 2011] Shevchenko, P. V. (2011). *Modelling Operational Risk Using Bayesian Inference*. Springer.

[Shevchenko and Del Moral, 2015] Shevchenko, P. V. and Del Moral, P. (2015). Valuation of barrier options using sequential monte carlo. *Journal of Computational Finance*.

[Shiryaev, 1996] Shiryaev, A. N. (1996). *Probability*. Springer-Verlag, New York.

[Siller, 2013] Siller, T. (2013). Measuring marginal risk contributions in credit portfolios. *Quantitative Finance*, 13(12):1915–1923.

[Skilling, 2006] Skilling, J. (2006). Nested sampling for general bayesian computation. *Bayesian analysis*, 1(4):833–859.

[Sklar, 1959] Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

[Sklar, 1996] Sklar, A. (1996). Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*, pages 1–14.

[Sofronov, 2013] Sofronov, G. (2013). An optimal sequential procedure for a multiple selling problem with independent observations. *European Journal of Operational Research*, 225(2):332–336.

[Targino et al., 2015] Targino, R. S., Peters, G. W., and Shevchenko, P. V. (2015). Sequential Monte Carlo samplers for capital allocation under copula-dependent risk models. *Insurance: Mathematics and Economics*, 61:206–226.

[Targino et al., 2016] Targino, R. S., Peters, G. W., Sofronov, G., and Shevchenko, P. V. (2016). Optimal exercise strategies for operational risk insurance via multiple stopping times. *Methodology and Computing in Applied Probability*, pages 1–32.

[Tarullo, 2008] Tarullo, D. K. (2008). *Banking on Basel: the future of international financial regulation*. Peterson Institute.

[Tasche, 1999] Tasche, D. (1999). Risk contributions and performance measurement. *Report of the Lehrstuhl für mathematische Statistik, TU München*.

[Tasche, 2008] Tasche, D. (2008). Capital allocation to business units and sub-portfolios: the Euler principle. In *Pillar II in the New Basel Accord: The Challenge of Economic Capital*, pages 423–453. Risk Books.

[Tran et al., 2014] Tran, M.-N., Scharth, M., Pitt, M. K., and Kohn, R. (2014). Importance sampling squared for bayesian inference in latent variable models. *Available at SSRN 2386371*.

[Tsanakas and Millossovich, 2016] Tsanakas, A. and Millossovich, P. (2016). Sensitivity analysis using risk measures. *Risk Analysis*, 36(1):30–48.

[Tweedie, 1957] Tweedie, M. (1957). Statistical properties of inverse gaussian distributions. *The Annals of Mathematical Statistics*, 28(2):362–377.

[Vergé et al., 2015] Vergé, C., Dubarry, C., Del Moral, P., and Moulines, E. (2015). On parallel implementation of sequential Monte Carlo methods: the island particle model. *Statistics and Computing*, 25(2):243–260.

[Vergé et al., 2016] Vergé, C., Morio, J., and Del Moral, P. (2016). An island particle algorithm for rare event analysis. *Reliability Engineering & System Safety*, 149:63–75.

[Villén-Altamirano and Villén-Altamirano, 1991] Villén-Altamirano, M. and Villén-Altamirano, J. (1991). RESTART: A method for accelerating rare event simulations. *Analysis*, 3:3.

[Wang et al., 1997] Wang, S. S., Young, V. R., and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and economics*, 21(2):173–183.

[Watson, 1922] Watson, G. (1922). *A treatise on the theory of Bessel functions*. Cambridge University Press.

[Willmot and Lin, 2001] Willmot, G. E. and Lin, X. S. (2001). *Lundberg approximations for compound distributions with insurance applications*, volume 156. Springer Science & Business Media.

[Wills, 2016] Wills, S. (2016). ORX response to the BCBS on the SMA. Online: `https://www.orx.org/Lists/PublicDocuments/ORX%20Response%20to%20the%20BCBS%20on%20the%20SMA.pdf`, Accessed: 13/Jul/2016.

[Wüthrich, 2015] Wüthrich, M. V. (2015). Non-Life Insurance: Mathematics & Statistics. *Available at SSRN 2319328, version of 29/Jun/2015*.

[Wüthrich and Merz, 2015] Wüthrich, M. V. and Merz, M. (2015). Stochastic claims reserving manual: Advances in dynamic modeling. *Swiss Finance Institute Research Paper*, (15-34).